Defining genomic drivers of evolution in the entomopathogenic *Serratia* spp.

A thesis
submitted in partial fulfilment
of the requirements for the Degree of
Doctor of Philosophy

at
Lincoln University
by
Amy Louise Vaughan

Lincoln University
2021

Abstract of a thesis submitted in partial fulfilment of the requirements for the Degree of Doctor of Philosophy.

Defining genomic drivers of evolution in the entomopathogenic *Serratia* spp.

by

Amy Louise Vaughan

There is a growing demand for new and sustainable solutions as alternatives to chemical pesticides both in Aotearoa New Zealand and worldwide. Since the early 1990s, isolates of the bacterium *Serratia entomophila* have been utilised as biocontrol agents, causing amber disease specific to larvae of the New Zealand endemic grass pest, *Costelytra giveni* (Coleoptera. Scarabaeidae). *Serratia proteamaculans,* another causal agent of amber disease, is as yet unutilised in commercial biocontrol but has been implicated in causing diverse disease pathotypes in grass grub larvae and larvae of other endemic scarabs (*Pyronota* spp.). As some isolates have a broader host range than *S. entomophila*, there is potential for *S. proteamaculans* to be a more suitable candidate for future biocontrol approaches.

Virulence determinants for both chronic and hypervirulent isolates reside on variants of a megaplasmid, pADAP (amber-disease associated plasmid), which encodes the Afp (antifeeding prophage) and Sep-Toxin complex. Pathogenic isolates of both species are found coexisting with non-pathogenic conspecifics (predicted to be around 44% of the *S. entomophila* population) in sampled field sites around New Zealand. It is yet to be elucidated why the pADAP plasmid is only found in a proportion of the population, if this is a stable population, or what benefit can be conferred by the presence or absence of the plasmid. There is potential that disease expression and retention of pADAP in the population is in some way mediated by the chromosome. This study seeks to further investigate the potential role of the chromosome of *Serratia* spp. in the population dynamic and variability of disease and plasmid persistence.

The presence of conspecific plasmid-free isolates compounds the need for characterisation of the production of accessory virulence determinants in entomopathogenic *Serratia*. Through assessment of 52 whole chromosomal sequences, it was found that while *S. entomophila* pADAP bearing isolates shared a relatively conserved chromosome, that of *S. proteamaculans* were heterogeneous. The consistent and predictable disease phenotype caused by *S. entomophila* isolates could therefore infer that *S. entomophila* may favour a chronic disease state. Pangenome wide analysis of *S. entomophila* isolates found differences in those from areas devoid of grass grub, suggesting the native pADAP

carrying isolates have undergone speciation, coevolving with *C. giveni* larvae. Comparative genomics showed chromosomal inversions and larger chromosome size in isolates from areas where amber disease has not previously been documented. Evidence of horizontal gene transfer (HGT) was found in both *S. entomophila* and *S. proteamaculans*, though predictions of pangenome bounds and putative islands suggest that these events are more commonplace in *S. proteamaculans*.

Plasmid transconjugants to non-native backgrounds were constructed to investigate the potential coevolution of the plasmid and chromosome. Differential expression of accessory virulence determinants between transconjugants and wildtype isolates demonstrated an association between down-regulation of various factors and acquisition of a non-native plasmid type. *In vivo* analysis supported this finding, where transconjugants with non-native plasmids displayed a more erratic disease phenotype. *In silico* assessment found chronic isolates to encode a similar number of accessory virulence determinants to their hypervirulent counterparts, suggesting a higher regulatory mechanism could be responsible for the decrease in time to infectivity found in disease caused by these isolates. Overall, the expression of accessory virulence determinants was found to be similar between pathotypes. Differences were exclusive to extracellular nuclease production (in non-pathogenic isolates and *S. entomophila* vs *S. proteamaculans*) and siderophore production in hypervirulent isolates.

While the plasmid disease-encoding genes are well characterised, the mechanism driving retention of chromosomally located accessory virulence determinants such as phospholipases, chitinases, and extracellular proteases in S*. entomophila* and *S. proteamaculans* are yet to be described*.* Bioinformatics and both *in vitro* (plate assays) and *in vivo* bioassays were used to assess accessory virulence factors (associated with host tissue damage and colonisation) in *Serratia* spp. isolates. Results were correlated with geographic/phenotypic data to determine any evolutionary links to places of isolation or virulence. Overall, this study has shown that the pADAP plasmid *of S. entomophila* and *S. proteamaculans* have coevolved, where regulatory mechanisms of chromosomally bound accessory virulence determinants potentially mediate pathogenic adhesion and host tissue degradation. Though evidence currently indicates that the *S. entomophila* genome is not reducing, it is hypothesised that the indicators of reduced HGT opportunity and evidence of trophic specialisation could suggest coevolution with *C. giveni*.

Keywords: Chromosome, pADAP, plasmid, horizontal gene transfer, entomopathogen, pangenome, comparative genomics, accessory virulence determinants, transcriptomics

Acknowledgements

I would like to thank the following people for their contributions to this study and my personal development throughout.

I would first and foremost like to thank my supervisors, Dr. Mark Hurst, Prof. Travis Glare, Dr. Karen Armstrong, Prof. Peter Fineran, Prof. Murray Cox, and Dr. Paul Gardener. I am extremely grateful for the long hours of guidance and expertise at often unreasonable hours. I have appreciated the skills, experience, and know-how you have guided me through from experiments to writing.

I would also like to thank my lab group, especially Amy Beattie and Mitchell Weston for their support, advice, and technical skills throughout my project. In addition, I would like to thank my fellow PhD colleagues. Marion and Lesley, whose knowledge and friendship has helped me through the last few years. Connor, thank you for all the cakes, they've been greatly appreciated.

Dr. Chikako Van Koten, for taking the time to explain and teach me statistical analysis. I've valued every one of our conversations.

Charles Hefer, for assistance in RNASeq transcriptome analysis. I am grateful for your speedy analysis and detailed explanations. I would also like to thank Dr. Aurelie Laugraud, Dr. Darrell Lizamore, and Ruy Jauregui for their skills and expertise in bioinformatics and comparative genomics. Who without, I would still be fruitless trying to understand how to unpack a tarball. I appreciate the time you spent teaching me the basics step by step.

A throwback but no less important. One of my most overdue thanks goes to Sarah Kellagher, who without her unwavering belief in me eleven years ago, I would not have got into science in the first place. Thank you for never quitting on me.

I would also like to thank my family. My parents, Gary and Nikki, and my siblings; Laura, Hollie, and Jack, whose support and encouragement, yet lack of sympathy, got me through this whole thing.

To my New Zealand family. My Abuelos and Abuelas. I love you all and you all know who you are. I have to make a special mention for Mariona, Carla, and Majo. I owe you all Carl Lasagans.

I would finally like to thank my partner, Caleb. Who, throughout this whole escapade, has been a 'pebble' for me to lean on and my biggest cheerleader. I could not have done this without your unwavering support, kindness, and persistence in keeping me on a schedule.

Table of Contents

Abstract ..... **iii**

Acknowledgements ..... **v**

Table of Contents ..... **vi**

List of Tables ..... **xi**

List of Figures ..... **xiv**

List of Abbreviations ..... **xviii**

Chapter 1 Introduction ..... **- 1 -**

1.1 New Zealand agriculture ..... - 1 -

The New Zealand grass grub *Costelytra giveni* ..... - 1 -

1.1 Pest control ..... - 3 -

1.1.1 Agricultural chemicals ..... - 3 -

1.1.2 Biological control ..... - 4 -

1.2 *Serratia* spp. ..... - 5 -

1.2 *Serratia* spp. diseases in grass grub ..... - 6 -

1.3 Ecology of amber disease, what drives pathogenicity? ..... - 10 -

1.4 Symbionts to pathogens ..... - 11 -

1.5 Drivers of evolutionary change in pathogen and commensals ..... - 12 -

1.5.1 Genome rearrangement of the chromosomal regions ..... - 12 -

1.5.2 Horizontal gene transfer ..... - 13 -

1.5.3 Phages ..... - 14 -

1.6 Plasmids ..... - 14 -

1.6.1 Plasmids and disease ..... - 16 -

1.6.2 Chromosomal effects on plasmid maintenance ..... - 16 -

1.7 Regulation of disease ..... - 17 -

1.8 Pathogen associated genes ..... - 18 -

1.8.1 Proteases ..... - 18 -

1.8.2 Lipases ..... - 19 -

1.8.3 Chitinases ..... - 19 -

1.8.4 Chitin binding proteins ..... - 19 -

1.8.5 Siderophores ..... - 20 -

1.9 Research relevance ..... - 21 -

1.10 Hypotheses ..... - 22 -

1.11 Objectives ..... - 23 -

Chapter 2 General Methods **24**

2.1 Bacterial strains and plasmids 24

2.2 Isolates of *Serratia* spp. investigated 25

2.3 Microbiological methods 27

2.3.1 Cell growth 27

2.3.2 Plate assays 27

2.3.3 Indole-3-acetic acid expression 29

2.3.4 Megaplasmid Visualization 30

2.3.5 Curing of plasmids 30

2.3.6 Growth curves 31
2.3.7 Microscopy 31
2.3.8 Mitomycin C Induction protocol 31
2.4 Molecular protocols 32
2.4.1 Genomic DNA sample preparation 32
2.4.2 RNA sample preparation 32
2.4.3 Plasmid mini-preparation method 33
2.4.4 BOX A1R PCR 33
2.4.5 Standard PCR protocol 34
2.4.6 Colony PCR 35
2.4.7 Fusion PCR 36
2.4.8 Restriction enzyme-based mutagenesis 37
2.4.9 Ligation 37
2.4.10 Chemical transformation 38
2.4.11 Electro transformation 38
2.4.12 Restriction digestion for mutation 38
2.4.13 Ethanol precipitation 38
2.4.14 Chromosomal and plasmid tags 39
2.4.15 Conjugation 39
2.4.16 Validation and stability of plasmid transconjugants 40
2.4.17 Sequence validation 40
2.5 Bioinformatic analysis 40
2.5.1 PacBio sequencing assembly 41
2.5.2 Illumina sequencing assembly 41
2.5.3 Removal of plasmid sequences from genomes 41
2.5.4 Annotation of assembled sequences 41
2.5.5 Average nucleotide Identity 42
2.5.6 BRIG alignment 42
2.5.7 Mauve 42
2.5.8 Roary 43
2.5.9 Pangenome analysis 43
2.5.10 Scoary 43
2.5.11 Breseq 44
2.5.12 Islandviewer4 44
2.5.13 Genome mining 44
2.5.14 COGs 45
2.5.15 16S ribosomal DNA (rDNA) phylogenies 46
2.5.16 Core genome phylogenies 46
2.5.17 RNA transcriptome analysis 47
2.6 Assaying virulence 47
2.6.1 Grass grub collection 47
2.6.2 Bioassay 47
2.6.3 Dose-response bioassay 48
2.6.4 Enumeration of bacteria from larvae 48
2.6.5 Pathotype determination of bioassays 49
2.6.6 Protozoa lysate protocol 49
2.7 Statistical analysis 50
2.7.1 Statistical analysis of the plate assays 50
2.7.2 Statistical analysis of bioassay data 50

Chapter 3 Delineating species boundaries of *Serratia* spp. 51
3.1 Introduction 51

3.2 Results 52
3.3 Discussion 59

Chapter 4 Bacterial phenotyping of the *Serratia* spp. 61
4.1 Introduction 61
4.2 Results 62
4.2.1 *Serratia* spp. isolates 62
4.2.2 Bioassays of wildtype isolates against grass grub larvae 64
4.2.3 BOX-fingerprinting 68
4.2.4 Plasmid visualisation in *S. entomophila* 73
4.2.5 Screening for accessory virulence determinants 76
4.2.6 Screening for traits for plant growth-promotion 84
4.3 Discussion 87

Chapter 5 Defining variation in the *Serratia entomophila* genome 91
5.1 Introduction 91
5.2 Results 92
5.2.1 *Serratia entomophila* isolate 626 92
5.2.2 Genome statistics for *S. entomophila* 98
5.2.3 MLSA phylogeny 102
5.2.4 Genome-wide analysis 105
5.2.5 The openness of the *S. entomophila* genome 109
5.2.6 Breseq 112
5.2.7 Mauve 113
5.2.8 Targeted mutagenesis of a phage-like cluster on the *S. entomophila* chromosome 120
5.2.9 Acquisition of foreign DNA 121
5.2.10 Accessory virulence factors in *S. entomophila* 128
5.2.11 Itaconate mutant 137
5.3 Discussion 148

Chapter 6 Comparative genomic analysis of isolates of *Serratia entomophila* and *Serratia proteamaculans* 155
6.1 Introduction 155
6.2 Results 156
6.2.1 Composition and comparison of *S. proteamaculans* genomes 156
6.2.2 Full core phylogenies of *S. entomophila* and *S. proteamaculans* 162
6.2.3 BRIGs comparison analysis 165
6.2.4 Determining the presence of advantageous metabolic pathways 170
6.2.5 Chromosomal comparisons 175
6.2.6 Detection of genomic Islands and regions of gene acquisition 182
6.2.7 Pangenome wide association and core genome comparative analysis 191
6.2.8 The openness of the *S. proteamaculans* pangenome 195
6.2.9 Chromosomally encoded accessory virulence determinants 196
6.2.10 Prediction of biosynthetic clusters 200
6.3 Discussion 205

Chapter 7 Chromosomal regulation of fitness in *Serratia* 209
7.1 Introduction 209
7.2 Results 211

7.2.1 MitC induced expression of virulence factors 211
7.2.2 Conjugation of plasmid types into various chromosomal background 220
7.2.3 LC_{50} and LT_{50} against grass grub larvae 223
7.2.4 Bacterial infectivity 229
7.2.5 Visual cell differences of transconjugants 231
7.2.6 Assessment of transconjugant growth relative to wildtype 233
7.2.7 Accessory determinants and plant beneficial traits 236
7.3 Discussion 244

Chapter 8 Transcriptional regulation of non-native plasmid transconjugants 250
8.1 Introduction 250
8.2 Results 251
8.2.1 Regulation of genes of interest 251
8.2.2 Plasmid gene regulation 259
8.3 Discussion 262

Chapter 9 General Discussion 265
9.1 Is *S. entomophila* genetically conversed compared to *S. proteamaculans*? 265
9.1.1 Is *S. entomophila* undergoing genome reduction? 265
9.1.2 Species-specific gene acquisition 266
9.2 Do metabolic clusters in *S. entomophila* aid maintenance of a chronic disease state? 268
9.3 Do chromosomally encoded genes have a role in disease? 269
9.4 Where is disease evolution heading, and what is the role of HGT? 270
9.5 Do chromosomes select for plasmids? 271
9.6 Is there a geographic link to strain occurrence? 273
9.7 Phylogenetic implications of this study 273
9.8 Conclusion 274
9.9 Future work 274
9.9.1 Genomic comparisons of *S. entomophila* and *S. proteamaculans* 274
9.9.2 Itaconate degradation cluster as a metabolically beneficial system 275
9.9.3 Geographic diversity of *C. giveni* genotypes 275

References 276

Appendix A Methods 291
A.1 Growth medium for IAA auxin detection assay 291
A.2 Kado and Liu megaplasmid visualisation buffers 291
A.3 Plasmid mini prep solutions 291
A.4 M9 Minimal Salts medium recipe 292

Appendix B Chapter 4 293
B.1 Comprehensive pathotype and geographic isolate data 293
B.2 Bioassay data 297
B.3 Full plate assay results 297

Appendix C Chapter 5 297
C.1 ANI matrix including *Serratia ficaria* isolate 457 297
C.2 Roary output file 297

C.3 Scoary output files 297
C.4 Comparative genome statistics including *Serratia ficaria* 457 298
C.5 Core genome alignment of only *Serratia entomophila* 300
C.6 Annotation of 40 Kb region isolate 626 301
C.7 Annotation of 33 Kb region isolate 477 304
C.8 Genomic island predictions of *Serratia entomophila* 307
C.9 Accessory virulence factor motif prediction 310

Appendix D Chapter 6 311
D.1 BRIGS diagrams including isolate keys 311
D.2 Roary output file 312
D.3 Scoary output files 313
D.4 Full KofamKOALA results tables 313
D.5 Genomic island predications of *Serratia proteamaculans* 317
D.6 Merops compiled presence absence 320

Appendix E Chapter 7 321
E.1 Putative SOS binding sit predictions 321
E.2 Transconjugant plate assay data 321

Appendix F Chapter 8 322
F.1 Transcriptome data files 322
F.2 pADAP annotation in comparison between A1M02 and 477 + pA1M02 322

List of Tables

Table 1 Traditional pesticide classifications and example pesticidal derivative products.....- 3 -
Table 2 Host bacterial strains and plasmid vectors used in this study.....24
Table 3 Antibiotics and growth substrates used in this study.....25
Table 4 Isolates of *Serratia* used in this study. Isolates were obtained from the AgResearch Culture Collection (Lincoln New Zealand).....26
Table 5 Standard PCR protocol reagent mix for a 50 µL reaction.....34
Table 6 Primers used in this study Sequence restriction sites are shown underlined where present.....34
Table 7 List of isolates identified for specific 16S rRNA region sequencing in addition to the original genome sequence.....46
Table 8 Isolates defined into each grouping as defined by 16S rRNA phylogenies Isolates shown are isolates to be further used in *in vitro* and *in silico* comparative analysis.....57
Table 9 Day 12-day bioassay assessment of selected *Serratia entomophila* and *Serratia proteamaculans* isolates.....65
Table 10 Isolates of *Serratia* spp. and their assigned pathotype as determined in this study..... 67
Table 11 Mega plasmid visualisation of 56 isolates of *Serratia entomophila* both plasmid (+) and non-plasmid (-), and their relative size.....73
Table 12 COG breakdown of *Serratia entomophila* 626 in comparison to five type isolates of species of *Serratia* from GenBank.....96
Table 13 Summary genome statistics of individual *Serratia entomophila* isolates assessed in the study.....100
Table 14 BLASTp of the highest associated genes with putative pathogenicity traits using Scoary. 106
Table 15 IslandViewer4 hits predicted in *Serratia entomophila* isolate 626.....124
Table 16 Presence of predicted transposases found on the chromosome of *Serratia entomophila* isolates using ISfinder displayed as a visual alignment.....127
Table 17 HMMER results for chitinase glyco-hydro family motif search and chitin-binding protein family on *S. entomophila* isolate 626, with percentage similarity of protein sequence to the most similar protein amino acid sequence on UniProt.....129
Table 18 HMMER results for lipase family motif searches on *Serratia entomophila* isolate 626, with percentage similarity to most similar protein on UniProt.....131
Table 19 Identification of extracellular secreted nucleases *Serratia entomophila* isolate 626, and its closest % similarity from UniProt BlastP function.....132
Table 20 Extracellular serine proteases and metalloproteases predicted in *Serratia entomophila* isolate 626 AA sequences BLASTp against Uniprot database for closest related protein homolog.....134
Table 21 AntiSMASH analysis for predicted secondary metabolite cluster in all isolates of *Serratia entomophila*. AntiSMASH results compiled into a presence or absence table for each predicted secondary metabolite cluster. Annotations are as shown fromantiSMASH output.....136
Table 22 Closest BLASTX hits to genes located in the *Serratia entomophila* itaconate degradation operon and associated transcriptional regulator.....139
Table 23 Virulence for mutant Mcl1:tet and the wildtype control of *Serratia entomophila* isolate 626 to grass grub larvae, including LC_{50} and LT_{50} with standard error calculation P values (Fisher's exact), with statistical significance in relation to the negative control, highlighted in bold.....144
Table 24 Genome statistics of individual *Serratia proteamaculans* isolates.....158
Table 25 Genome statistics for additional *Serratia* spp. and *Yersinia* used in this study.....161
Table 26 Number of complete modules identified by KofamKOALA for metabolic units located on chromosomes of selected *Serratia proteamaculans* and *Serratia entomophila*174

Table 27 IslandViewer4 hits predicted *in Serratia proteamaculans* hypervirulent isolate AGR96X 183
Table 28 Predicted transposases found on the chromosome of *Serratia proteamaculans* isolates using ISfinder 186
Table 29 Prediction of CRISPRcas arrays in *Serratia* isolates Number of arrays encoded per chromosome indicated in brackets. 187
Table 30 Phage predictions of intact and incomplete phage regions in all isolates Predictions made using Phaster. 189
Table 31 HMMER results for chitinase glyco-hydro family motif search and chitin-binding protein family on *Serratia proteamaculans* isolate AGR96X with percentage similarity of protein sequence to the most similar protein amino acid sequence on UniProt. 197
Table 32 HMMER results for lipase family motif searches on *Serratia proteamaculans* isolate AGR96X, with percentage similarity to the most similar protein on UniProt 199
Table 33 AntiSMASH results of the hypervirulent isolate AGR96X 201
Table 34 AntiSMASH predicted secondary metabolite clusters in *Serratia proteamaculans S. liquefaciens* isolates are bolded. 203
Table 35 Transconjugants used in this study 210
Table 36 Average disease and mortality rates of Mitomycin C induced isolates of *Serratia proteamaculans* and *Serratia entomophila* 214
Table 37 Motif search results of CTGT(N_8)ACAG on various genomes of *Serratia* isolates used in MitC induction. 216
Table 38 Predicted genes associated with putative SOS binding sites in *Serratia proteamaculans* AGR96X SOS binding sites identified using the gammaproteobacterial motif CTGT(N_8)ACAG and the Geneious motif search function. 217
Table 39 Isolates with varying pathotypes and plasmids encoded selected for tarns conjugant experiment 220
Table 40 Successfully transconjugated bacterial chromosomes with the inserted plasmid and the corresponding chromosomal and plasmid markers 221
Table 41 Day 12 percentage disease based bioassay data for all transconjugants and the wildtype controls, including LC_{50} and LT_{50} with standard error calculation P values (Fisher's exact), with significance in relation to the negative control, highlighted in bold 226
Table 42 Log2fold change observed for regulator, *nucC*, associated with *nucA* expression in relation to 255
Table 43 SepABC genes differential expression between wildtype A1M02 plasmid donor and transconjugant 477 + pA1M02. 260
Table 44 *Afp* genes differential expression between transconjugant 477 + pA1M02 and wildtype A1M02 plasmid donor. 261

Table B 1 Geographic location data for all isolates used in this study Location is to the closest known point. Virulence associated with results from this and previous studies. Sequenced refers to sequencing via either Illumina or PacBio further used in comparative genomics in this study 293

Table C 1 BlastP using Uniprot of the 40Kb chromosomal region unique to *Serratia entomophila* isolate 626 302
Table C 2 BlastP using Uniprot of the 33Kb chromosomal region unique to *Serratia entomophila* isolate 477 305

Table D 1 Number of complete modules identified by KofamKOALA for metabolic units located on chromosomes of additional *Serratia entomophila* 314
Table D 2 Number of complete modules identified by KofamKOALA for metabolic units located on chromosomes of selected *Serratia proteamaculans* and *Serratia entomophila*315
Table D 3 Number of complete modules identified by KofamKOALA for metabolic units located on chromosomes of selected *Serratia proteamaculans*.....316

Table F 1 Transcriptome data of pADAP genes in wildtype A1M02 (MW 7-9) and transconjugant 477 + pA1M02 (AV 7-9).....323

List of Figures

Figure 1 New Zealand agriculture land use ANZ 2008 land use map acquired from Ministry for the Environment (MFE). 1
Figure 2 *Costelytra giveni* grass grub Demonstrates early 3rd instar grub located within the top 2-5cm of the soil. 2
Figure 3 Damage to grass root system caused by *Costelytra giveni* (Left) shows *Costelytra giveni* larva found in the roots of grass circled in red, and (right) shows the 'lifting' effect on the grass and top layer of soil after damage caused by grass grub larvae. 3
Figure 4 Annotation of the *Serratia entomophila* A1MO2 pADAP plasmid as described by Hurst et al. (2011). 7
Figure 5 Three step fusion PCR diagram shown 36
Figure 6 Schematic of restriction site targeted mutagenesis. 37
Figure 7 16S rDNA Maximum likelihood tree of sequenced *Serratia* spp. 52
Figure 8 16S rDNA Maximum likelihood tree of sequenced *Serratia entomophila* and *Serratia proteamaculans* in this study 53
Figure 9 16S rDNA phylogeny for *Serratia entomophila* 55
Figure 10 16S rDNA phylogeny for *Serratia* isolates and type strains with final classifications. 58
Figure 11 Geographic distribution *of Serratia* spp. used in this study across New Zealand ...63
Figure 12 *Serratia entomophila* plasmid free isogenic isolate 5.6 compared with WT plasmid carrying *A1M02 Serratia entomophila* AGR96X used as a control. 69
Figure 13 Bacterial BOX fingerprinting of *Serratia entomophila* 70
Figure 14 Geographical isolation of each isolate and their corresponding BOX fingerprinting profiles 1-3 three BOX profiles shown with colour coordination in the key. 71
Figure 15 BOX bacterial fingerprinting of Chatham Islands isolates compared to. 72
Figure 16 Kado and Liu mega plasmid visualisation of all isolates of *Serratia entomophila* using isolate A1M02 as a reference for the 153 Kb mega plasmid 74
Figure 17 Summarised response of all isolates to plate assays for six treatments, and auxin detection 96 well assay of *Serratia* 77
Figure 18 Results of protease plate assays across all *Serratia entomophila* and *Serratia proteamaculans* isolates 79
Figure 19 Results of lipase plate assays across all *Serratia entomophila* and *Serratia proteamaculans* isolates 80
Figure 20 Results of chitin plate assays across all *Serratia entomophila* and *Serratia proteamaculans* isolates 81
Figure 21 Results of siderophore plate assays across all *Serratia entomophila* and *Serratia proteamaculans* isolates 82
Figure 22 Results of DNase plate assays across all *Serratia entomophila* and *Serratia proteamaculans* isolates 83
Figure 23 Results of auxin 96 well assays across all *Serratia entomophila* and *Serratia proteamaculans* isolates 85
Figure 24 Results of phosphatase plate assays across all *Serratia entomophila* and *Serratia proteamaculans* isolates 86
Figure 25 Map of sequence homology from assembly to reference genome. 93
Figure 26 Genome atlas for *Serratia entomophila* 626. The outermost circle shows BlastP similarities against non-redundant databases. 94
Figure 27 COG distribution comparisons for *Serratia entomophila* and type strains of *Serratia* spp. 97
Figure 28 ANI values of all *the Serratia entomophila* isolates displayed in a heat map adapted from matrix from ANI/AAI Matrix. 102

Figure 29 MLSA Phylogenetic analysis A) MLSA phylogenetic analysis of *Serratia entomophila* isolates with *Serratia ficaria* 457 (indicated with *) and *Serratia ficaria* type strain NCTC 12148 analysed by maximum likelihood. 104
Figure 30 Comparative genome statistics of the core genome of *Serratia entomophila* defined by Roary 107
Figure 31 Core genome alignment of all chromosomes of *Serratia entomophila* and *Serratia ficaria*. 108
Figure 32 Core genome reduction analysis of sequenced *Serratia entomophila* chromosomes from the core genome compiled with Roary pangenome software. 109
Figure 33 Pangenome analysis made in pangenome plots utilising Roary output. 110
Figure 34 OrthoVenn representation of predicted proteins shared between representatives of each monophyletic group. 112
Figure 35 Mauve alignment of *Serratia entomophila* isolates 626 (top) and 477 (bottom). 115
Figure 36 Mauve alignment of multiple sequenced isolates. Comparison of various isolates of geographic isolation and pathotype of *Serratia entomophila*. 116
Figure 37 Large ~90 Kb region in Chatham Island isolate 440 that is not present in West Coast isolate MC2. 118
Figure 38 Mauve alignment of isolates *Serratia entomophila* 626 (top) and 440, and *Serratia ficaria* 457 (bottom) 119
Figure 39 Flanked phage-like region found in *Serratia entomophila* 626 120
Figure 40 Induced expression of virulence factors in *Serratia entomophila* 626 and DinI::spec mutant. 121
Figure 41 Predicted genomic islands in *Serratia entomophila* 123
Figure 42 Predicted genomic islands in *Serratia entomophila* isolate 440 125
Figure 43 Schematic of the *Serratia entomophila* itaconate degradation pathway operon. 138
Figure 44 Predicted degradation pathway utilised in *Serratia entomophila*, modified from (Sasikaran et al. 2014) and MetaCyc pathway for itaconate degradation. 138
Figure 45 Chromosomal orientations of the degradation operon A) 158, 176, 210, 219, 345, 398, 440, 442, 477, 626, A1M02, I, iDIA, MC2, Man3 B) 220, 305, 364,482, 1343 140
Figure 46 Phylogeny of Mch amnio acid sequence of the itaconate degradation cluster in *Serratia entomophila* with protein homologs AA sequence phylogeny. 141
Figure 47 Growth curves of WT 626, Mcl1::tet and complemented *mcl1* gene in optimal and stress conditions 48 h growth curves in triplicate with standard error shown. 142
Figure 48 *In vivo* competitive growth experiment. 145
Figure 49 Protozoa (indicated by black arrow refer inset for expanded view) in a 1:10 dilution of larval fermentation sac pooling sample. 146
Figure 50 Distribution of genomic G+C content of isolates included in this study P-value calculated by Kruskal Wallis P-value 159
Figure 51 Average nucleotide identity of *Serratia proteamaculans* isolates (%). 160
Figure 52 ANI values of all isolates assessed in this study Isolates clustered to similar ANI values. 162
Figure 53 Core genome phylogeny of closest related *Serratia* isolates. 164
Figure 54 BRIG diagrams of both *Serratia entomophila* and *Serratia proteamaculans*. 167
Figure 55 Alignment of hypervirulent AGR96X to other isolates of *Serratia proteamaculans* 169
Figure 56 Alignment of *Serratia entomophila* isolates to Chatham Island isolate 440 Innermost rings from additional Chatham Island isolates (442 and 294) and West Coast isolate MC2. 170
Figure 57 Shared protein clusters in representative isolates of each species of *Serratia* spp. sequenced. 175
Figure 58 Mauve whole chromosome alignments of two isolates of *Serratia entomophila* (626 and 345) alongside four isolates of *Serratia proteamaculans* (1048, 142, 143) Similarity skew is shown in collinear blocks. 177

Figure 59 Variable region encoding metalloprotease on *Serratia entomophila* 178
Figure 60 Overall comparison of the chromosomes of *Serratia proteamaculans* AGR96X (top) and *Serratia entomophila* 626 (bottom) highlighting points of interest 180
Figure 61 Itaconate region of *Serratia entomophila* missing from *Serratia proteamaculans* AGR96X 181
Figure 62 Predicted genomic islands on the *Serratia proteamaculans* chromosome *of Serratia proteamaculans* subsp*. quinovora* 184
Figure 63 Pangenome of isolates of *Serratia proteamaculans* 193
Figure 64 Presence and absence graph of genes in all sequenced isolates 194
Figure 65 Rarefaction curve of *Serratia proteamaculans* pangenome gene cluster analysis compared with *Serratia entomophila* made in Micropan 195
Figure 66 Rarefaction curve *of Serratia proteamaculans* subsp. *proteamaculans* pangenome gene cluster analysis made in Micropan 196
Figure 67 Schematic of predicted siderophore/aryl polyene region in *Serratia proteamaculans* AntiSMASH prediction of siderophore aryl polyene biosynthetic cluster in most isolates of *S. proteamaculans*. 202
Figure 68 Presence/ absence of identifiable accessory virulence factors and biosynthetic clusters in *Serratia proteamaculans* and *Serratia entomophila* 204
Figure 69 MitC bioassay against grass grub larvae, 2018-2020. 212
Figure 70 Percentage larvae exhibiting disease phenotype after bacterial SOS response induction with MitC using culture supernatant without ultracentrifugation over six days. 215
Figure 71 Kado and Liu megaplasmid visualisation and BOX PCR validation of transconjugants. 222
Figure 72 Sample bioassay disease instances of wildtype isolates at day six for wildtype controls 227
Figure 73 Sample bioassay disease instances of transconjugants at day seven. 228
Figure 74 Number of CFUs recovered from host macerates of larvae inoculated with transconjugants in relation to their relative wildtype strains over 15 days and normalised data represented as $\log_{10}$ transformed percentage of initial CFU. 230
Figure 75 Plate morphology and microscopy of 477 and transconjugant 477 + pAfpX after 24 h growth. 232
Figure 76 Plate morphology and microscopy of 3041 and transconjugant 3041 + pAfpX after 24 h growth 232
Figure 77 Growth curve of transconjugants and their wildtype counterparts 234
Figure 78 Growth curves of transconjugants and their wildtype counterparts 235
Figure 79 Response of all conjugates and their wildtype chromosomal counterparts to accessory virulence detection and plant growth promotion assays. 237
Figure 80 DNase expression across all transconjugant isolates and their wildtype counterparts. 239
Figure 81 Protease expression across all transconjugant isolates and their wildtype counterparts. 240
Figure 82 Chitinase degradation assays across all transconjugant isolates and their wildtype counterparts. 241
Figure 83 Lipase degradation assays across all transconjugant isolates and their wildtype counterparts. 242
Figure 84 Phosphatase solubilisation assays across all transconjugant isolates and their wildtypes. 243
Figure 85 IAA (auxin) detection assays (µg/mL) across all transconjugant isolates and their wildtype counter parts. 244
Figure 86 Analysis of transcriptome data in triplicate comprising samples AV1-3 (WT 477) and AV7-9 (477 + pA1M02) 252
Figure 87 Principal component analysis of in triplicate comprising samples AV1-3 (WT 477) and AV4-6 (477 + pAfpX) 253

Figure 88 Transcriptome reads count of *nucA* nuclease in transconjugants 477 + pAfpX and 477 + pA1M02 compared to wildtype 477. 254
Figure 89 Transcriptome reads count of chitinase genes in transconjugants 477 + pAfpX and 477 + pA1M02 compared to wildtypes 477 and A1M02. 255
Figure 90 Transcriptome reads count of chitin binding proteins in transconjugants 477 + pAfpX and 477 + pA1M02 compared to wildtypes 477 and A1M02. 256
Figure 91 Transcriptome reads count of phospholipase genes in transconjugants 477 + pAfpX and 477 + pA1M02 compared to wildtypes 477 and A1M02. 257
Figure 92 Transcriptome reads count of metalloprotease and extracellular proteases genes in transconjugants 477 + pAfpX and 477 + pA1M02 compared to wildtypes 477 and A1M02. 258

Figure 93 Analysis of transcriptome data in triplicate comprising samples MW7-9 (WT A1M02) and AV7-9 (477 + pA1M02) A) 259
Figure C 1 Comparative genome statistics of the core genome of *Serratia entomophila* with *Serratia ficaria* isolate 457, defined by Roary 299
Figure C 2 Core genome alignment of all chromosomes of *Serratia entomophila* 300
Figure C 3 Predicted genomic islands in isolates of *Serratia entomophila* 308
Figure C 4 Predicted genomic islands in isolates of *Serratia entomophila* 309

Figure D 1 BRIGs diagram with isolate key of *Serratia proteamaculans* 311
Figure D 2 BRIGs diagram with isolate key of *Serratia entomophila* using Chatham Islands isolate 440 as a reference 312
Figure D 3 Predicted genomic islands in isolates of *Serratia proteamaculans* 318
Figure D 4 Predicted genomic islands in isolates of *Serratia proteamaculans* 319

List of Abbreviations

Abbreviation	Definition
Afp	Anti-feeding prophage
AfpX	Anti-feeding prophage X- hypervirulent encoding
ALA	aminolevulinic acid
Amp	Ampicillin
ANI	Average nucleotide identity
β gal	β galactosidase
BLASTn	Nucleotide sequence/ query BLAST search
BLASTp	Protein sequence/ query BLAST search
bp	Base pair
CFU	Colony-forming units
Cm	Chloramphenicol
COG	Cluster of Orthologous Group
DNA	Deoxyribonucleic acid
GI	Genomic island
HGT	Horizontal gene transfer
Kb	Kilobase
KEGG	Kyoto Encyclopaedia of Genes and Genomes
LB	Luria-Bertani growth medium
LD_{50}	Lethal-dose 50. Minimum dose to cause 50% death
LT_{50}	Lethal-time 50: minimum time to cause 50% death
Mb	Mega base
MitC	Mitomycin C
MLSA	Multi Locus Sequence Analysis
NCBI	National Center for Biotechnology Information
OD	Optical density
pADAP	Amder disease-associated plasmid
PCR	Polymerase chain reaction
RNA	Ribonucleic acid
RNA-Seq	Whole transcriptome shotgun sequencing
rpm	Rotations per minute
Sep-TC	*Serratia entomophila* pathogenicity toxin complex
Spec	Spectinomycin
SNP	Single nucleotide polymorphism
TC	Toxin complex
Tet	Tetracycline
tRNA	Transfer-ribonucleic acid
X-gal	5-bromo-4-chloro-3-indolyl-β-D-galacto-pyranoside

Chapter 1

Introduction

1.1 New Zealand agriculture

Aotearoa New Zealand's agriculture system is based on a primarily pastoral model, with 32% agricultural land and dairy around 9.8%. As of 2018, approximately 40%, or 10.7 million hectares, of New Zealand's total landmass was covered with exotic grasses such as rye and clover for grazing (Ministry for the Environment and Stats NZ 2018) (Figure 1). These numbers are decreasing, with the largest land-use changes showing a change in exotic grass pastures of approximately -1.7% per annum. Overall, the continued decrease in agricultural land use in pastoral farming land aligns with high country tenure changes which have moved high country grasslands into the Department of Conservation's control (Ministry for the Environment and Stats NZ 2018). Grazing pastures, however, remain the backbone of New Zealand's economy (Ballingall and Lattimore 2004).

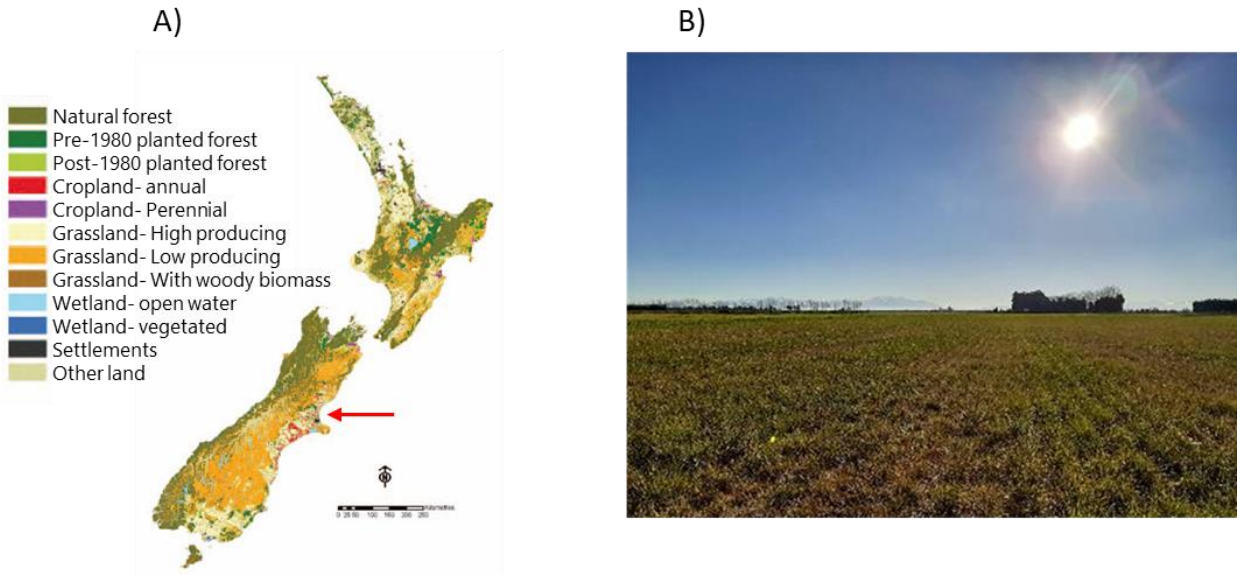


Figure 1 New Zealand agriculture land use
ANZ 2008 land use map acquired from Ministry for the Environment (MFE). Land use allocations displayed in the key. B) An example of Canterbury grazing pasture grasslands. Red arrow indicates the location of the Canterbury plains.

The New Zealand grass grub *Costelytra giveni*

The New Zealand grass grub, *Costelytra giveni* (Coleoptera: Scarabaeidae), formerly *C. zealandica* (Coca-Abia and Romero-Samper 2016) and commonly referred to in the larval stage as grass grub, is an endemic scarab larva that has transitioned from native tussock to pasture pest (Figure 2). The

species can also be an issue in horticultural, fruit, and cereal crops across both Te Ika ā Māui (the North Island) and Te Wai Pounamu (the South Island) of New Zealand. There are three stages of larval growth between egg to pupa. First instar larvae feed immediately on root systems from hatching in November-December, with second instar larvae from January to February occurring in the top 6 cm of soil. The third instar, occurring from late February to June at usually 2-5 cm depth in the soil (though they can be found as deep as 20 cm), causes the most damage to crop root systems. Pasture infestations typically remain in localised areas due to immediate mating on the emergence of mature females, with greater spread enabled through strong winds (Townsend 2002). Larvae prepare for pupating through increasing size and build-up of fat reserves through the prolonged 3rd instar stage when gut contents are voided. Between late October to early December, the adult scarab beetles emerge (Bain 1980), although this varies with climate. Colder, high country plots under adverse environmental conditions such as drought or cold have been shown to go through two yearly cycles rather than the yearly cycle observed in the pastures of low land areas such as Canterbury and Otago.

Figure 2 *Costelytra giveni* grass grub
Demonstrates early 3rd instar grub located within the top 2-5 cm of the soil.

With the introduction of non-endemic crops to New Zealand, the larval stages of *C. giveni* successfully adapted to an altered diet from natural tussock to the root systems of ryegrasses and white clover (Lefort et al. 2015). Population densities in excess of 66 larvae/m^2 were sufficient to cause severe damage to pasture. Large densities of grass grub can decimate these root systems

leading to identifiable browning of grasses and the 'lifting' of the top layer of soil when pulled (Figure 3). Extensive damage calls for an effective and rapid method of pest control.

Figure 3 Damage to grass root system caused by *Costelytra giveni*
(Left) shows *Costelytra giveni* larva found in the roots of grass circled in red, and (right) shows the 'lifting' effect on the grass and top layer of soil after damage caused by grass grub larvae.

1.1 Pest control

1.1.1 Agricultural chemicals

For millennia modern humans have been utilising materials for their insecticidal properties, such as ash for protection against fleas and lice in ancient civilisations (Hakbijl 2002). In modern agriculture, chemicals have been a longstanding part of arable crop and pasture grazing methods of farming as ways to reduce damage caused by pest insects. These chemical insecticides can be classified into six synthetic organic groups (Table 1).

Table 1 Traditional pesticide classifications and example pesticidal derivative products

Chemical composite grouping	Pesticidal derivative
Organochlorines	DDT, BHC
Organophosphates	Malathion, fenthion, dichlorvos
Carbamates	Carbaryl, bendiocarb
Pyrethroids	Deltamethrin, permethrin
Neonicotinoids	Imidacloprid, thiamethoxam
Ryanoids	Chlorantraniliprole

Organochlorines, such as dichlorodiphenyltrichloroethane (DDT) operate via pathways regulating sodium channels in myelinated nerve fibres, causing prolonged exposure of the nerve to the Na^+ current (Vijverberg et al. 1982). After environmental reporting in the mid-20th century, these chemicals were commercially banned from agricultural use, due to their environmental and carcinogenic effects

(Conis 2010). Bioaccumulation of organochlorine pesticides in aquatic systems results in lethal and sublethal effects of both wildlife and human biota via food chain accumulation and pollution of waterways (Chopra et al. 2011). Pyrethroids have been shown to follow a similar mode of action on these gated channels (Vijverberg et al. 1982). Organophosphates are used extensively in insect control targeting the insect nervous system via enzymatic interference of cholinesterase (Fukuto 1990). These have high levels of toxicity in humans and wildlife, prevalent in water, especially areas where the continuous discharge of pesticides into water flows occurs (Sandoval-Herrera et al. 2019). Fat-soluble chemicals such as fenthion have a symptomatic onset of up to 12 h post-ingestion and due to their affinity for storage fat- reserves can take days to eliminate (Kwong 2002). Neonicotinoids have lower toxicity to mammalian life as a synthetic analogue of nicotine and a longer persistence after field application. These act by activating acetylcholine receptors in the insect's nervous system inducing tremors, paralysis, and death. As broad-spectrum insecticides, they have drawn criticism and have been cited as a factor in the decline of honeybees, leading to a ban in seed dressings across the EU imposed in 2013 (Blacquière and van der Steen 2017, Woodcock et al. 2018, Chambers et al. 2019). Alongside cultural control including rolling pastures, insecticides are utilised in the control of the previously mentioned *C. giveni* larvae (Townsend 2002). Diazinon has been a standard insecticide for grass grub control in New Zealand, though mandated deregistration from 2028 in New Zealand (Environmental Protection Authority 2013) and increased concerns over short efficacy have pushed for changes (Zydenbos et al. 2016). Zydenbos et al. (2016) found that after two years post application larval populations in Diazinon-treated plots, grass grub re-established to higher concentrations than that of control plots.

Following the results of these studies, a consumer-driven push has increased pressure to reduce the application of traditional chemical pesticides for a more sustainable and beneficial approach to pest control. New methods should have potential ecological benefits for wildlife, the environment, and pose no issues for human health (Nicolopoulou-Stamati et al. 2016, Blacquière and van der Steen 2017).

1.1.2 Biological control

Biological control agents (BCAs) are non-chemical alternatives to traditional pesticides, providing a method for pest control avoiding effects to non-target native flora and fauna. BCA's can be developed from naturally occurring entomopathogenic viruses, protozoans, bacteria, and fungi alongside other insectivorous species (Köhl et al. 2019). Biological control methods have been used to control pest damage in crops for centuries, though the development of synthetic pesticides in the mid-20th century all but eliminated use (Barratt et al. 2018). Due to increasing consumer pressure for organic, residue-free foods, not treated with traditional chemical pesticides, the growing global trend in demand for

biological controls continues to increase yearly (Bailey et al. 2010). Their mode of action as of targeted, specified host control, as well as having a low environmental burden, make them a more preferable and sustainable option (Barratt et al. 2018). To be effective, BCAs generally need to establish within a niche in the target environment, where factors such as competitive soil biota and non-favourable conditions do not hinder their persistence and resultant efficacy (Ojiambo and Scherm, 2006).

Currently, microbial BCA's are developed commercially globally, including in New Zealand. The bacterium *Bacillus thuringiensis* is widely used in pesticides, accounting for approximately 65% of biopesticide usage worldwide. As a highly specific BCA, *B. thuringiensis* provides effective control for some pest species, namely caterpillars (Palma 2017). As a broad range microbial control, *Brevibacillus laterosporus* has recorded toxicogenic activity on various invertebrates, including parasitic nematodes (Ruiu 2013). Two New Zealand-specific insect diseases are caused by *Serratia* spp. and *Yersinia entomophaga,* which were isolated from diseased grass grub larvae and have potential as BCA's (Jackson et al. 2001, Hurst et al. 2011, Chattopadhyay et al. 2017). These *Serratia* species are host-specific to *C. giveni*, a single species of endemic Scarabaeidae (Jackson 1991). *Yersinia entomophaga* has been described to have a broad-host-range insect pathogen, including for control of Scarabaeidae, with the highest control among insect species of Lepidoptera and Coleoptera (Hurst et al. 2011). Application of *Serratia* spp. as a specific BCA has been described as more effective in comparison to chemical control with diazinon for grass grub (Zydenbos et al. 2016).

1.2 *Serratia* spp.

The current documented species of *Serratia* spp. are members of the Yersiniaceae family of gram-negative, bacteria, consisting of opportunistic pathogens of both plants and eukaryotes. Species of *Serratia* have been isolated from a diverse array of habitats, and some are known pathogens of humans, plants, insects, coral, and nematodes. Five species of the genus have been isolated from water, suggesting that this habitat could be the primary habitat of the genus (Grimont 2006). While documented, rare human infections caused by *Serratia* spp. are primarily caused by *S. marcescens,* though infections through *S. liquefaciens S. plymuthica, S. quinivorans, S. grimesii, S. fonticola,* and *S. ficaria* have also been recorded (Mahlen 2011). Many species of *Serratia* encode secondary metabolites that promote the synthesis of compounds that confer a selective advantage, potentially including antimicrobial properties that benefit bacterial survival in the soil.

Although the presence of insecticidal toxins in *Serratia* spp. are documented, *Serratia entomophila* and *Serratia proteamaculans,* the model organisms used in this study, appear to be truly entomopathogenic.

1.2 *Serratia* spp. diseases in grass grub

Amber disease was first described in Canterbury by Trought et al. (1982) as a soil-borne infection. named for the amber-like discolouration (from the usual grey appearance) in the larval gut associated with the accumulation of infection. The disease manifests as a cessation of feeding and gut clearance of organic matter. A combined effect of weakness and sepsis result in the death of the larvae. It was initially thought that disease was caused by the soil-dwelling bacterium *Hafnia alvei,* which was later reclassified as *S. entomophila* (Trought et al. 1982, Grimont et al. 1988). As previously discussed, *Serratia entomophila* and *Serratia proteamaculans* are the causal agents of amber disease in *C. giveni*. Amber disease is a chronic infection

The pathology of amber disease displays as high cell densities within the larval midgut- but haemocoel invasion is not observed until 1-3 months after initial infection (Grimont et al. 1988). As yet, the mechanisms utilised by *Serratia* spp., including the causal agents of amber disease, to cross the haemocoel is currently undefined. Observations in other bacterial species show mechanisms have developed to directly invade the haemocoel of the host. *Photorhabdus luminescens* and *Xenorhabdus nematophila* are examples of bacteria that have evolved to be assisted in transport via nematodes, and breach the gut epithelia or cuticle to then deliver bacteria directly into the haemocoel (Vallet-Gely et al. 2008). Utilisation of toxin haemolysins by *Xenorhabdus* to overcome host defences and incapacitate the immune system facilitates pathogen entry to the haemocoel resulting in sepsis and host death (Hurst 2016). Type III secretion systems (T3SS) are required in *Salmonella* to mediate internalisation by translocating effectors onto the host epithelial cell membrane, targeting macrophages for pyroptosis before colonising a niche region to establish infection in mammalian cell lines (Ramos-Morales 2012). This mechanism is likely utilised by *Y. entomophaga* for rapid haemocoel invasion (Hurst et al. 2015). Accessory virulence determinants such as degradative enzymes, for example, chitinases, can play a role in the interaction between host and pathogen. *Xanthomonas* spp. secretes degradative enzymes through the type II secretion system (T2SS) to enable bacterial interactions with plant host cells (Büttner and Bonas 2010). Chitinases target chitin, a polysaccharide found in epithelial gut linings of insects, and can evoke a pesticidal response when administered through a BCA. The *Yersinia entomophaga* toxin complex (Yen-TC) and its associated chitinases of *Y. entomophaga* result in the dissolution of the peritrophic membrane, causing damage and larval death of the diamondback moth *Plutella xylostella* (Landsberg et al. 2011).

*Serratia entomophila a*nd *S. proteamaculans* are relatively distant species from each other based on prior 16S analysis, and therefore to cause the same phenotypic disease in the same insect host species they must share similar mechanisms for virulence (Glare et al. 1993, Grkovic et al. 1995). Through

assessment of plasmid profiles, a similar-sized plasmid was found to reside in pathogenic isolates of both *S. entomophila* and *S. proteamaculans*, designated amber disease-associated plasmid (pADAP) (Glare et al. 1993). It was postulated that the plasmid has been acquired through horizontal gene transfer (HGT) (conjugation) between bacterial species. Two of the required virulence gene regions are encoded in pADAP (Figure 4). One, the Antifeeding prophage (Afp), is a contractile virus-like particle that, once activated, causes rapid cessation of feeding in larval challenges 2-5 days post-ingestion. The translated products of the Afp encoding region form a similar structure to bacteriophage tails and pyocins and have been termed tailocins. A three-day LD_{50} of approximately 512 ± 3.1 Afp particles was determined by Rybakova et al. (2013) to cause death in larval challenges (Hurst et al. 2004, 2007). The second region, termed Sep (*S. entomophila* pathogenicity), is an insect toxin complex, translated from three genes, *sepA*, *sepB,* and *sepC*, that mitigates clearance of the midgut and causes the characteristic amber coloration (Jackson et al. 1993, Hurst et al. 2007).

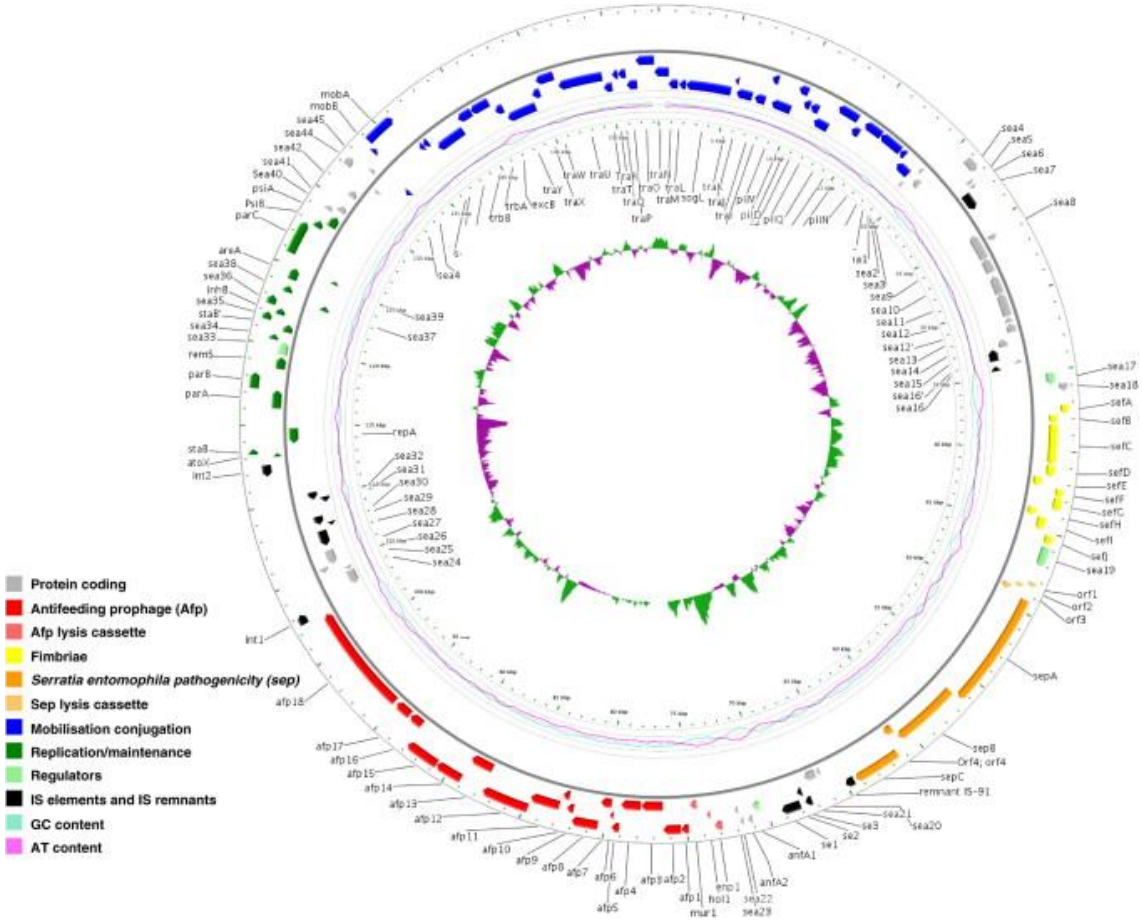


Figure 4 Annotation of the *Serratia entomophila* A1MO2 pADAP plasmid as described by Hurst et al. (2011)
Required virulence regions are highlighted, Afp (red) and sep (orange). Inner histogram denotes G+C skew plotted as sequence deviation, where negative (purple) and positive (green).

Constant exposure to the *sepABC* orthologues to maintain a diseased state is presumably characteristic of the chronic nature of *S. entomophila* induced disease- where death can occur up to three months post-infection. The provision of Sep proteins elicits symptoms of amber disease in larvae of *C. giveni*, whereas reversion to healthy grub occurs within 14 days if the supply of proteins is ceased. This suggests that a constant supply of Sep proteins is required to maintain disease (Hurst et al. 2007). To account for the slow disease progression, it is thought that the production of both Afp and Sep is likely tightly regulated (Sitter 2020). Through the co-expression of Afp and Sep, cessation of feeding and gut clearance leads to a depletion of larval fat reserves. This then leads to gut weakening and eventual bacterial breach of the haemocoel, causing septicaemia and death. Optimal growth of *S. entomophila* occurs in the midgut, though as previously mentioned no specific site of colonisation asides from food particles have been identified (Jackson et al. 2001).

Recent research by Sitter (2020) assessing the plasmidome of *Serratia* spp. found that the plasmids represented diverse structures despite being encoded by a singular genus. It was hypothesised that specialisation to a specific niche promoted genetic diversity within pADAP, ergo adaptation to the larval host. Although the plasmids are highly stable, and no incompatibility grouping was found with the chromosomes of *Serratia* spp., niche evolution of the host cell restricted HGT events, therefore, increasing plasmid diversity. An example of this diversity is the identification of a hypervirulent isolate of *S. proteamaculans* by Hurst et al. (2018). Disease phenotypes caused by this hypervirulent isolate were characterised by invasion of the haemocoel resulting in rapid death within 5-12 days post-challenge. Genomic sequencing of the hypervirulent isolate identified a variation of Afp, designated AfpX, which encoded additional toxin components and two Afp16 tail-length termination proteins. The AfpX variant from AGR96X pADAP has functional pathogenicity against larvae of both *C. giveni* and the New Zealand manuka beetle *Pyronota festiva*, which is a more diverse host range than the Afp characterising disease of *S. entomophila* (Hurst et al. 2018). Identification of this novel AfpX prophage variant revealed that further variants of amber disease-associated virulence determinants could be isolated and may confer a biological advantage to survival in particular host niches.

Serratia entomophila and *S. proteamaculans* have been isolated from soil in high cell numbers (up to 10^5 CFU/g soil) where natural epizootic events have occurred, though typically represents only a small proportion of the natural soil microflora (Jackson 1990, Jackson 1999). Populations of *S. entomophila* have, however, also been found to decline in the absence of grass grub larvae, implicating a relationship between the two species. Environmental sampling for the presence of *S. entomophila* across different New Zealand soil types indicates that the bacterium is restricted to survival in broad soil types, or the grass grub hindgut (O'Callaghan 1989). Soil isolates of *Serratia* spp. have been

differentiated through various plate media selective method types to determine species identification. Historically, itaconate selective medium has been used as a way to differentiate isolates of *S. entomophila* from other *Serratia* spp. (including *S. proteamaculans*) due to the species unique ability to degrade itaconate as a carbon source (Grimont et al. 1988, O'Callaghan 1993). Conversely, *S. proteamaculans* is globally more ubiquitous -and found in further environmental niches including plants, fungi, and decaying plant material (Grimont et al. 1981). This bacterium is therefore likely to have a greater opportunity to make contact with other *Serratia* spp., increasing the bacterium's opportunity to acquire genetic material (Yang et al. 2016). Variable efficacy of *Serratia* spp. as successful BCA's can only be recorded when the ability of the microbe to function competently in the environment of the pest has been assessed. The period necessary for *S. entomophila* to establish control of grass grub can be as long as 2-3 years, after which there is typically a grass grub population collapse at around 5 years (Grimont et al. 1988). This form of control is significantly slower than the deployment of a synthetic pesticide where death typically occurs within 2-3 weeks of application. Post application of *S. entomophila*, an annual increase of 15- 20% in disease levels is a preliminary indicator of success, demonstrating the prolonged time required for full pest control (Jackson 1999).

In 1990*, S. entomophila* isolate A1 (later renamed A1M02) was commercialised as a biocontrol agent under the name Invade® (Jackson 1990). The original Invade® formulation was a refrigerated liquid containing approximately 4.0×10^{10} cells/mL. This needed to be seed drilled (buried at a specific and consistent depth) into the soil to be effective. Invade® was plagued with variability issues, resulting in inefficient coverage in treated pastures including bacterial sedimentation in long-term storage. Substantial viability loss of the bacterium over time in cold storage was observed (Pearson 1995). Further assessment reduced the predicted shelf-life (with optimum bacterial survivability) from 20 weeks to seven days of storage, which became a cost versus effectiveness issue. Re-development was necessary to resolve the issues with the original formula (Johnson et al. 2001). More recently, Invade® was rebranded and reformulated as Bioshield™ with the introduction of an inert granular formulation and the use of *S. entomophila* isolate 626 to replace A1M02 (Johnson et al. 2001). To be an effective formulation, Bioshield™ needed to initiate an epizootic event to spread amber disease through the larval population. This was made possible as a stable granule allowed for a targeted, higher efficacy application alongside shelf-life longevity. Application of Bioshield™ has been shown to increase rates of amber-disease in established grass grub populations by approximately 20% more than in untreated pastures with natural epizootic events (Johnson 2004). In addition to increased field efficicacy, long-term shelf life issues were resolved with a bacterial stability of approximately six months in granular form at room temperature compared to the previous seven days. A wider geographic market could therefore be targeted and pastures could be treated 'as is' and not as an on demand market. Final

improvements were then made to enhance field trials, replacing isolate A1M02 for the more consistent and viable isolate 626.

1.3 Ecology of amber disease, what drives pathogenicity?

The ecological niche of *Serratia* was undefined for many years due to its confusing status and early misidentifications. The first *Serratia* spp. to be identified, *Serratia liquefaciens*, was initially identified as *Aerobacter liquefaciens* and was not reclassified until 1971 after extensive comparisons with *Serratia marcescens* (Bascomb et al. 1971). Many *Serratia* are regularly isolated from insect hosts, with isolations of *S. proteamaculans, S. marcescens,* and *S. liquefaciens* well reported (Grimont et al. 1979). The scarcity of reports of infection with some species, namely *Serratia rubidaea*, has been attributed to the species lack of chitinase production (a virulence factor broadly associated with the infection of insect hosts) to enable it to break through the chitinous insect gut (Lysenko 1976). Overall, the *Serratia* genus colonises a broad range of ecological niches not limited to the soil and microbiota of insect guts.

Some isolates of plasmid-containing *S. entomophila* and *S. proteamaculans* result in a chronic infection resulting in the typical amber disease phenotype. Analysis undertaken by Dodd et al. (2006) and Glare et al. (1993) found that approximately half of all isolates under assessment collected from pasture areas were amber-disease causing, which led to a general assumption that approximately 50% of field samples are pathogenic whilst the remaining 50% non-pathogenic conspecifics. More recently some isolates, such as the prior mentioned AGR96X, have been identified to have a hypervirulent pathotype resulting in a far shortened infection rate (Hurst et al. 2018).

Based on the observed homogeneity of *S. entomophila*, one of the hypotheses for the minor variation found between isolates of *S. entomophila* is specific environmental adaptation according to geographic location. Findings by Dodd (2003) revealed that only two variants of pADAP were identified from *S. entomophila* when comparing the DraI plasmid restriction profiles. The lack of variation in these plasmids suggests either later evolution of *S. entomophila,* limited microbial contact for HGT, or that an evolutionary bottleneck previously occurred. A reduced genome, as a typically stable structure, could result in the tight regulation of the disease process wherein *S. entomophila* may be coevolving with grass grub. *Serratia proteamaculans* isolates contained more highly varied plasmids (Sitter 2020), possibly resulting from insertion elements and the acquisition of multiple different virulence-associated regions (Hurst personal communication). The greater variety in genomic profile seen with *S. proteamaculans* could be indicative that this bacterium has evolved through chromosome encoded mechanisms to enable it to reside in a specific ecological niche where it then acts as an opportunistic pathogen (Dodd 2003).

1.4 Symbionts to pathogens

For the evolution of a commensal organism to a pathogen, the bacteria must (1) adapt to fully colonise a host, (2) circumvent the host immune system, (3) facilitate replication, (4) propagate, and (5) compete with resident microflora (antimicrobials). This is mitigated by a variety of accessory virulence factors of which some of these will be discussed later, and underlying mechanisms to subvert the host resources for bacterial gain (Alberts 2002).

However, due to the lack of co-adaption and the decreasing benefits from causing death to a host, there was a school of thought that commensalism is the evolved stable state and that pathogenicity traits are effectively down-regulated as the species evolves, though this theory is generally now disregarded (Brown et al. 2006). Some bacteria have developed from commensalism (close coexistence) to become symbiotic (mutually beneficial), leading to host dependency by undergoing genome alterations and often a reduction in the number of genes. Selective pressures resulting in beneficial trait selection of syphoning nutrition from the host, or competing with other bacterial colonies, can later facilitate the induction of pathogenic traits (Ehrlich et al. 2008). Although these mechanisms are clear, the advantages of acquiring such genes and the evolutionary benefits are poorly understood.

Endosymbionts share common features of reduced genomes when compared to that of free-living organisms, but to different degrees. Increased host-to-bacterial interaction can result in the pathogen obtaining secondary metabolites directly. This effectively makes these beneficial biological pathways redundant, increasing the ability for the bacteria to sustain a symbiotic existence, but decreasing a bacterium's ability to survive without the host. This has been described in species of micro-bacteria such as *Rickettsia* and Mycoplasma that are completely reliant on host interactions as a source of energy (Moran 2002). Genome reduction can also lead to a loss in the function of acquiring foreign genomic material that is disadvantageous to bacterial evolution, leading to a mostly clonal population (Kurokawa et al. 2016). Conversely, genome reduction can also prove to be beneficial to bacterial fitness. This has been demonstrated in a laboratory trial with a twenty percent reduction of the genome in a strain of *Bacillus subtilis* (MBG874) using a series of multiple deletion mutants which increased expression of several other gene products when compared to its wildtype and benefited host cell metabolism (Morimoto et al. 2008). Naturally coevolved host-specific associations demonstrating reductions in beneficial genomes have also been described. For example, high levels of stable genome associations can result in an increased loss of redundant pseudogenes, as can be observed in the *Blockmanni*a coevolved symbionts of ants (Williams and Wernegreen 2015). The level of reduction can be varied depending on the stage of reduction or the level of coevolution the organisms have reached. *Serratia symbiotica* displays variable levels of genome reduction under

comparative genomic assessment. *Serratia symbiotica* isolate SAf shows hallmarks of genome reduction and compaction when compared to free-living *S. marcescens* isolate DB11, whereas isolates SCt and progressively more so isolate ST display further levels of genome reduction. This leads to metabolic dependencies with widescale gene loss (Manzano-Marín and Latorre 2016). These patterns make for an identifiable model of genome reduction in the *Serratia* genus that could potentially apply to other *Serratia* species. If *S. entomophila* is indeed moving to a symbiotic relationship, this might explain the chronic nature of the disease in grass grub caused by this pathogen, whereby selection for virulence genes prolongs disease effects and ultimately prolongs the death of its host.

1.5 Drivers of evolutionary change in pathogen and commensals

Mutation, as described by (Hershberg 2015), is the main driver of evolution. Through randomised events, point mutations (local), and genomic rearrangements (global), evolution can drive change in bacteria. The persistence of advantageous mutations have been attributed to natural selection favouring advantageous nucleotide changes (Comeron 2014). There are several mechanisms outlined here by which this can occur, with implications for bacteria to become pathogenic.

1.5.1 Genome rearrangement of the chromosomal regions

Rearrangement of chromosome encoded genes can occur by a range of mechanisms, duplication, deletion, and inversion leading to genome variability and therefore increased genetic variation between strains (Darmon and Leach 2014). These rearrangement events have been suggested as an adaptation and repair mechanism by bacterial chromosomes (Roth et al. 1996). For example, *Shigella* strains and *Escherichia coli* are genetically close enough to be characterised as the same species, making their differentiation challenging (Devanga Ragupathi et al. 2017). The evolution of *Shigella flexneri* from *E. coli* was facilitated by an acquired virulence plasmid, as well as deletions of genes encoding lysine decarboxylase activity, detrimental to pathogenicity (Maurelli et al. 1998). In comparisons *of Shigella* strains and *E. coli,* it has also been noted that the deletion of the *fliF* operon that encodes for the flagellar in *E. coli* results in non-motile *Shigella* (Devanga Ragupathi et al. 2017). In addition to deletions and duplications, inversions are also implicated in the 'evolvability' of bacterial chromosomes (Merrikh and Merrikh 2018).

Through whole-genome analysis, chromosomal rearrangements can be observed in otherwise similar isolates. Though no significant difference in average nucleotide identity (ANI) was found between two isolates of *Yersinia entercolitia* (isolate 105.5R® and 8081), a large 2.7 Mb region of rearrangement between the two isolates corresponding to approximately half the size of the complete genome was identified by Wang et al. (2011). Synteny breakpoints were described as flanked by insertion sequence (IS) elements, hypothesised to contribute to structural rearrangements in both genomes.

1.5.2 Horizontal gene transfer

One of the drivers of pathogen evolution, horizontal gene transfer (HGT), is defined as the movement of DNA and genetic material between unrelated unicellular individuals. Therefore, this can occur between species and strains as well as unrelated populations. This is in contrast to vertically transferred genes with occur inter-generationally, between related organisms. HGT is recognised as an important factor in pathogen evolution, resulting in the enrichment of a bacterial species by the acquisition of genetic information from other bacteria allowing the utilisation of new ecological niches (van de Guchte 2017). HGT and mechanisms that select for spontaneous genome rearrangement and mutation are responsible for wide variations in bacterial diversity across many related species, although in many systems they are poorly understood.

Bacterial evolution through HGT can be mediated by mobile genetic elements (MGEs) such as plasmids (San Millan et al. 2016). Transfer of material is undertaken by either (I) transformation, (II) conjugation, or (III) transduction (Gyles and Boerlin 2014). Transformation is mediated by the uptake of free DNA from the environment via pores in the bacterial cell wall that are capable of transformation. The receipt and integration of exogenous ssDNA rely on conserved proteins to mediate uptake. One strand is internalised (3'-5') through the ComEC membrane protein and is bound by DprA. RecA then drives recombination by finding a homologous region on the chromosomal DNA to form a donor and recipient heteroduplex (Mortier-Barrière et al. 2007). Conjugation is one of the easiest methods of transferring genomic information from a donor to a recipient cell, with well-defined modes of action. On plasmids the pilus tip of the pili contacts the recipient cell forming a stabilised structure whereby a DNA transport pore connects the cytoplasm of the two cells. *OriT* of the plasmid DNA is then cleaved by a relaxase and a single strand of plasmid DNA is covalently mobilised to Mob protein that then interacts with the proteins at the transport pore (Yin and Stotzky 1997). Both the recipient and the donor cell then synthesise a complementation DNA strand for the plasmid. In the case of the entomopathogenic *Serratia,* the mobilisation mechanisms of pADAP plasmid acquisition have been defined, with replicate studies finding conjugation of the pADAP and variant pADAP plasmids mediates the expression of virulence-associated traits (Grkovic et al. 1995, Sitter 2020). Transduction mediated transfer will be discussed further in section 1.5.3.

1.5.2.1.1 *Transposable elements*

Transposable elements (TEs) and transposons are areas of the genetic sequence that can relocate between loci (Schaack et al. 2010). As the simplest of mobile genomic elements (MGE's) at <2.5 Kb, an insertion sequence encompasses an open reading frame (ORF) with a transposase and two repeated sequences. When two of these Insertional Elements (IEs) flank a genetic region a mobile transposon is

formed allowing the gene region to become mobile (Gyles and Boerlin 2014). Transposable elements (TE's) can form a conjugative bridge through extracellular conjugative pili from host to recipient bacterial strain and facilitate the transfer of plasmids via HGT (Ochman et al. 2000). Through high sequence identity IS elements mediate genomic rearrangements. As TE's can only move limited amounts of DNA (e.g. individual genes), a plasmid or other mobile element is often involved in the transfer of larger amounts of genomic material from one bacterium to another.

1.5.2.1.2 *Genomic islands*

Genomic islands (GIs) are defined as areas of DNA, between 10 and 200 Kb in size, within a bacterial genome where past mobility is recognised through the differences between closely related bacterial strains (Juhas et al. 2009). Closely related GIs have been identified on both chromosomes and plasmids and are often flanked by IS elements allowing mobility (Gyles and Boerlin 2014). The increase of bacterial fitness resulting from the presence of a GI is likely positively selected for and is identified as a fitness island, of which four categories are named: (I) ecological islands, (II) saprophytic islands, (III) symbiosis islands, and (IV) pathogenicity islands (PAIs). Of relevance to this project, there are PAIs whose gene content directly affects the pathogenicity of the bacteria. Mobility of PAI varies resulting in diversity between virulence regions of closely related species. PAI's encoding superantigen toxins in *Staphylococcus aureus,* for example, transduce into new *S. aureus* strains easily and swiftly, resulting in a high diversity of toxin profiles across the species (Gyles and Boerlin 2014). Key characteristics of GI's infer their origin as from HGT events. tRNA's flanking a region encoding an integrase gene are often found as markers of a GI, in addition to altered G+C content relative to the whole genome (Ramsay et al. 2017). The loss of mobility genes over time allows the GI to then ameliorate the G+C content to that of the host cell, resulting in fixed integration through reductive evolution (Hayek 2013).

1.5.3 Phages

Transduction-mediated transfer of viral bacteriophage particles can have large effects on the acquisition of antibacterial or virulence-associated traits in lysogenic bacteria. Broudy et al. (2002) identified a secreted protein in *Streptococcus pyogenes* which was found to be a sequence homolog to streptococcal DNases carried by a lysogenic bacteriophage, thus inferring the importance of phages in the infection process. Not all phage-associated genes however are inducible. Loss of regulatory genes in integrated phages can render these mobile elements defective, losing the ability to express any virulence factor encoded in the region.

1.6 Plasmids

Plasmids, such as the aforementioned amber disease-causing pADAP, are circular units of DNA either conjugatively transferred by pili from one bacterium to another or inherited in linear evolution (vertical

gene transfer). Pili-mediated HGT is undertaken in the previously mentioned mechanism of conjugation. Conjugation is the only mechanism of the three HGT mechanisms to be mediated by cell-to-cell contact, utilised through plasmids transferred via conjugal pores (Frost et al. 2005). Conjugation via pilus (through which the movement of DNA is mainly facilitated) is mediated mostly by type 4 secretion system (T4SS) or pili involved in DNA transfer (called sex pili) (Filloux 2010). A group B type IV pilus has been identified as the mechanism of transfer for the PAPI-1 pathogenicity island in *Pseudomonas aeruginosa* (He et al. 2004).

Plasmids are projected to cause a range of fitness effects and burdens to their host bacteria, with the most beneficial plasmids hypothesised to eventually integrate into the host chromosome, while those posing the most burden could be lost over time. Plasmid loss has been characterised as a rare event. A study by (Glare et al. 1993) had to utilise significant effort through a heat curing mechanism to remove plasmids, involving high temperatures over long periods to obtain bacterial cells without the pADAP plasmid. Björkman et al. (2000) suggested that chromosomal compensation via mutation is a more favoured mechanism to adapt to costly plasmid types. As plasmids confer an 'exploitative' relationship on their host cell it could be assumed that bacterial cells would eventually submit to environmental selection pressures resulting in plasmid loss. Plasmid persistence is a less understood part of the plasmid puzzle. The ability of plasmids to persist in the environment and be positively selected for by bacterial communities is supported by the breadth of plasmid discovery in microbial niches. The persistence of these plasmid populations, though potentially carrying beneficial genes, can actively apply a fitness cost to the host that can result in the selection of non-plasmid carrying conspecifics. The ability for these plasmids to then persist within a population is referred to as the plasmid paradox (Harrison and Brockhurst 2012). The method utilised by plasmids to coadapt to their bacterial hosts has yet to be resolved, where beneficial genes should be ameliorated to the chromosome with redundant plasmids lost as an energy sink (Carroll and Wong 2018). However, plasmids have evolved to maintain their form. More recent experiments by Sitter (2020) found in both plate and flow cytometry studies that pADAP was not lost by its host cell even under high-stress conditions such as initiation of the SOS response and minimal medium, where a minimum amount of plasmid retention was noted at ~85%, showing high levels of plasmid stability. Conjugative plasmid stability can be impacted by the category of incompatibility groups (Inc). Classified based on the origin of replication (*ori*), genetic incompatibility for two plasmids to reside in the same chromosome (having identical replicons) is a regulatory pattern described widely in the Enterobacteriaceae (del Solar et al. 1998). Compatibility systems were described in the 1970s by Datta and Hedges (1971), with the inability of two plasmids to persist in a single cell line, ergo incompatible for coexistence. Exploration of chromosomal determinants that contribute to plasmid persistence is, therefore, an important contextual study for resolving the paradox.

1.6.1 Plasmids and disease

pADAP was first implicated in amber disease when megaplasmid visualisation identified a large plasmid of the same size in all but one pathogenic isolate of the *S. proteamaculans* and *S. entomophila* assessed (Glare et al. 1993). Glare et al. (1993) identified that heat curing the plasmid removed the pathogenicity of the isolate, therefore associating bacterial pathogenicity to the large plasmid found. The megaplasmid was further characterised for one isolate, A1M02 previously used as the BCA in control *of C. giveni* (Hurst et al. 2011). Further work by Dodd et al. (2006) then assessed further bacterial isolates for *sep* orthologs-genomic determinants of amber disease. The presence and variation of *sep*-like gene clusters on plasmids of *S. proteamaculans* indicated levels of mobility between the amber-disease genes on the pADAP plasmid.

Associations of toxins associated with disease to a plasmid complex are not new. *Clostridium perfingens* is known to carry up to three toxic plasmids each encoding up to three toxin-associated genes (Bannam et al. 2011). These *Clostridium* plasmid systems encode for a conserved 35 Kb transfer of clostridial plasmids (tcp) region that is implicated in the conjugative transfer of toxin plasmids, from infective to commensal isolates (Parsons et al. 2007). Regulation of these plasmid systems can be by two regulatory mechanisms. The VirS/VirR global regulator system (Cheung et al. 2009) and Agr-like quorum-sensing regulator (Ohtani et al. 2009). Of particular importance to biocontrol, the insecticidal endotoxins crystal (Cry) proteins of *Bacillus thuringiensis* (Bt) have been associated with large plasmids. Of the 770 different *cry* genes identified as of 2017, most were expressed on plasmids, allowing Bt to have novel combinations of insecticidal toxins and numerous applications in pest control (Rolle et al. 2005, Soares-da-Silva et al. 2015).

1.6.2 Chromosomal effects on plasmid maintenance

The maintenance of plasmids in daughter cells is integral to the maintenance of plasmid-encoded virulence regions and mobile gene elements, is mitigated by plasmid-specific-partitioning proteins. Defects and mutations in plasmid replication can allow for plasmid-free bacterial cells to proliferate without accessory genes that influence bacterial fitness and evolution (Hayes 2003). The integration of the plasmid into the chromosome can have benefits for bacterial fitness and pathogenicity regulation. For example, the ability of *S. flexneri* to impart disease depends on whether the virulence-related plasmid is integrated into the chromosome. Integration is associated with the downregulation of the expression of these genes and ensures plasmid maintenance (Zagaglia et al. 1991).

Post-segregational killing eliminates plasmid-free cells that arise due to replication errors often found in the host chromosome. These are instigated by toxin-antitoxin complexes (TA), of which there are six

described systems (Page and Peti 2016). These TA systems can also be encoded chromosomally, with evidence showing functionality as a stress response system, though this functionality hypothesis is being challenged (Saavedra De Bast et al. 2008). For instance, the presence of TA systems in *Vibrio chlolerae* has been implicated in hindering large-scale chromosomal deletions of a superintegron (Szekeres et al. 2007). *Escherichia coli* encodes for seven type II TA systems containing cassettes of two or three corresponding genes. TA cassettes are characterised by where an autoregulatory toxin-antitoxin loci overlap forms. Anti-toxin is regulated by the presence of a protease that rapidly degrades, whereas the plasmid presence then replenishes antitoxin rapidly. The toxin is then controlled by the presence of antitoxin- where replication not resulting in a plasmid renders production to be nil and therefore allows mitigation of the toxin to result in targeted cell death (Hayes 2003). The *yafNO* cassette in *E. coli* has been described previously as a TA system where the presence of *yafN* is only required in the presence of *yafO*.

1.7 Regulation of disease

Various mechanisms serve a purpose to bacterial toxin synthesis in pathogens. A cascade of regulation is usually in place, often regulated by two-component factor systems or global transcriptional regulators (Zhang et al. 2018). LysR-type transcriptional regulators (LTTRS) are ubiquitous amongst prokaryote organisms, as the most common regulatory protein. These regulators have been associated with the activation or repression of virulence factors. LcrX, a LTTS from *Xanthromas axonopodis* pv. glycines were demonstrated in proteomic analysis to be tied to the regulation of siderophore secretion. In *P. aeruginosa*, MvFR defective mutants showed an impaired degree of virulence in comparison to the wildtype strain (Reen et al. 2013).

Two-component regulatory systems (TCS) work to convert an environmental stimulus such as exposure to antibiotics, stress conditions, and nutrient availability into a phosphorylation or dephosphorylation response. In *Pseudomonas aeruginosa*, the GacS/GacA system is implicated in the regulation of virulence factors. This system comprises a sensor kinase (GacS) that transfers a phosphate group to the regulator (GacA), resulting in the upregulation of rRNAs (Rasamiravaka et al. 2015). GacS/GacA has been described as a 'super-regulator' of the Quorum sensing (QS) system, mediating the regulation of N-butyryl-L-homoserine lactone (Parkins et al. 2001). QS is a mechanism in bacteria that regulates the phenotypic expression of enzymes that can control virulence using cell-to-cell signaling. Regulation of accessory determinants such as proteases and lipases can potentially be attributed to the N-acyl-homoserine lactones(AHL) signaling autoinducer that accumulates in the environment in a cell density-dependent manner (Rutherford and Bassler 2012). Quorum sensing mechanisms work on the principle that at low cell densities, inducer concentrations are too low to be detected, whereas at higher cell densities the concentrations increase the global cell response. In amber disease caused by isolates of

S. entomophila, and unlike other *Serratia*, chitinases are thought to be down-regulated by Quorum sensing (Hurst, per comms). Induction of 2-fold of chitinolytic activities under addition of quorum sensor homoserine lactone (HSL) in *S. proteamaculans* B5a implies that the chitinases are upregulated. This also corroborates findings by Coulthurst et al. (2006) whereby mutation of the AHL synthesiser and subsequent supplementation with synthetic N-butanoyl-L-homoserine lactone (BHL) removed and restored the ability of the bacterium to produce chitinase (Christensen et al. 2003). This reduced chitinase expression at a high cell density is thought to prolong the persistence of *S. entomophila* in *C. giveni* (Hurst, 2016).

1.8 Pathogen associated genes

Through an array of accessory virulence factors, the chromosomes of pathogens have been shown to mediate disease through invasion, adhesion, and degradation of host cells and tissue. These mechanisms of disease are facilitated by a number of enzyme-encoding genes that are considered below. Therefore, the presence or absence of these genes can be an indicator of the types of niches they can occupy and the host-pathogen relationship.

1.8.1 Proteases

As ubiquitous enzymes, proteases are essential for life; irreversibly cleaving peptide bonds leading to downstream effectors on protein expression. Extracellular and serine proteases that interact with the host's cell surface are implicated as factors for the initiation of the infectious process in disease. Inactivation of the hosts' antimicrobial peptides and modification of host cell surface are just two of the potential ways in which serine proteases can be utilised by the pathogenic bacterium to evade the host immune system (Zdzalik et al. 2013). In gram-negative bacteria, secretion systems mediate the secretion of proteases, relying on proteolysis for multiple facets. Extracellular proteases are known to degrade host defence components, inhibiting the host response to bacterial invaders. Membrane and cytosol (Lon) bound proteases indirectly contribute to virulence by the determination of virulence regulators and bacterial tolerance to suboptimal environments (Frees et al. 2013). Vibriolysin, a metalloprotease from *Vibrio* species, has been implicated in such extracellular protease virulence by degrading most tissue via proteolysis, accelerating pathogenic infection (Miyoshi 2013). These mechanisms have been previously recorded in the *Serratia*, where many isolates exhibit proteolytic properties. Proteolytic strain SCBI (South African *Caenorhabditis briggsae)* was found to have highly transcribed levels of two serine proteases and two metalloproteases during the initial infection stage in entomopathogenic infection (Petersen and Tisa 2014). The identification of such virulence-mediated enzymes in amber disease causal agents could implicate chromosomal bound regions in the potential initial stages of infection, thus contributing to bacterial pathogenicity.

1.8.2 Lipases

Through cleavage of triglycerides, secreted lipases may play a direct role in early-stage infection by enhancing bacterial cell adhesion and causing direct damage to host cells and tissues (Stehr et al. 2003). Phospholipases in particular have been linked to triggering lysis of host cells affecting immediate damage (Ghannoum 2000). Specifically, extracellular phospholipases in *Candida albicans* have been found actively secreted in 79% of clinical isolates, with mutagenesis of the *plb1* gene encoding phospholipase B significantly attenuating the virulence potential of the organism in murine models (Park et al. 2013).

1.8.3 Chitinases

Family 18 and 19 glycohydrolase chitinases (EC.3.2.1.14) hydrolyse the polymer structure chitin; a major component of the epidermis, trachea, and peritrophic membrane of the gut epithelium in insects. Insect chitin therefore acts as a permeable barrier to the midgut (Merzendorfer and Zimoch 2003). Due to the relative abundance of this polymer as a key skeletal component in insects, the ability of a bacterium to degrade chitin can be a crucial factor in initiating infection. Chitinolytic action has been described in the entomopathogenic fungi *Beauveria bassiana*. Endochitinase (Bbchit1) overproduction resulted in enhanced virulence in aphids with a general reduction of LC_{50} by 50% (Fang et al. 2005). Glycoside hydrolase Family-18 and 19 chitinases encode the majority of chitinases, whereby family 18 generally includes prokaryotic and eukaryotic chitinases and family 19 significance lies with plant and nematodes (Kasprzewska 2003, Funkhouser and Aronson 2007). The soil-borne biocontrol agent *Serratia plymuthica* expressed chitinolytic activity via regulation of chitinase-associated genes similar to that of *S. liquefaciens* and *S. marcescens* ChiA. Regulation of activity was then implicated in the repression of spore germination in phytopathogenic fungi with up to 78% inhibition (Frankowski et al. 2001). The importance of these enzymes in biocontrol has, therefore, already been implicated in *Serratia* spp*.*, and potentially could be important in *S. entomophila* and *S. proteamaculans.*

1.8.4 Chitin binding proteins

In addition to chitinases, chitin-binding proteins (CBP's) have been associated with a bacteria's ability to bind to chitin aiding colonisation, as was described by Kirn et al. (2005), where a secreted CBP of *Vibrio cholerae* mediated attachment to epithelial chitin aiding infection. The localisation of a *Bacillus thuringiensis* binding protein (CBPA) on the bacterial cell surface was elucidated by Qin et al. (2020). Results implicated CBPA in the induction of *B. thuringiensis* cells in the midgut induced by an alkaline environment, facilitating adhesion to the chitin-rich peritrophic matrix, therefore, enhancing virulence performance of further virulence-associated enzymes. These results implicate chitin-binding proteins

in the early stages of entomopathogenesis, potentially becoming one of the primary binding responses to the colonisation of a chitin-rich gut epithelial tissue.

1.8.5 Siderophores

Bacterial requirements for iron (Fe) to facilitate their growth means they must adapt to be able to sequester iron from an already iron-limiting host via intrinsic mechanisms such as the secretion of iron-chelating siderophores (Sandy and Butler 2009). In environments prone to low iron presence, Fe^{2+} dependent repressors (*fur* gene) are inactivated, resulting in transcriptional activation and further secretion of siderophores into the environment to sequester iron. *Serratia marcescens* is a prime example of a siderophore secreting bacterium (Khilyas et al. 2016), where the production of catecholate siderophores correlated with exponential growth phase (Khilyas et al. 2016). Aerobactin synthesis has also been previously recorded in *Serratia liquefaciens* and *Serratia ficaria* (Angerer et al. 1992), with a wide distribution encoded across the Enterobacteriaceae. Aerobactin knockouts of hypervirulent *Klebsiella pneumoniae* (hvKP) resulted in a severe reduction in pathogenicity, displaying the relevance of this compound in human infection of hvKP (Khilyas et al. 2016). The implication of this aerobactin in virulence gives rise to the possibility of further Enterobacteriaceae having increased virulence based on iron-chelating compounds.

1.9 Research relevance

Identifying conserved and variable regions between related species of bacterial entomopathogens will improve understanding of bacterial evolution and the balance between virulent and avirulent isolates within an ecological niche. By determining if *S. entomophila* is truly clonal and its theorised capacity to only thrive within the gut of *C. giveni* we can understand what traits are unique in this New Zealand species. Underpinning the mechanisms of pathogen evolution by identifying the potential chromosomal drivers will elucidate the fitness or environmental benefits of bacteria to retain pathogenic tendencies. In addition, characterisation of the relationship between the chromosome and the virulence plasmid can be used as a model system in further investigations of plasmid persistence within populations.

The implications of this research could be extensive in the development of an effective biocontrol agent for grass grub in New Zealand. By defining the diversity of known *S. entomophila* and *S. proteamaculans* isolates and categorising these to known pathotypes, more efficient selection can be undertaken in the future by linking chromosomal fitness to environmental factors and pathotype. While genetic is not currently accepted in New Zealand, it is possible to select from environmental bacterial reservoirs for advantageous features. In this way, it is possible to determine whether, for example, the AGR96X hypervirulent isolate is the most suited and adapted isolate for commercial purposes, or whether the evidence indicates other isolates of *S. proteamaculans* afford more advantageous fitness determinants. The ability to accurately select relevant isolates in the search for more efficient pasture pest control would reduce the economic cost to the agriculture sector by millions of dollars every year. The development of environmentally sound biological controls over synthetic pesticides is becoming a more viable and sought-after alternative for the industry.

1.10 Hypotheses

This project investigates two *Serratia* species with known entomopathogenic characteristics found in New Zealand, isolated from disease larvae of *C. giveni.* Non-pathogenic conspecific isolates have also been isolated directly from the environment. While the plasmid pADAP and derivatives are known to encode the main disease determinants, it is unclear why some isolates of each species have the plasmid while others do not. Through the analysis of genomic sequences differentiation between 56 *Serratia* isolates and of the phenotype diversity among a cohort of 95 isolates, the mechanisms that mediate evolution and the adaptive advantages these bacteria have for entomopathogenesis may begin to be elucidated. To address this here, the following hypotheses will be tested:

i) ***S. entomophila*** **is a genetically conserved bacterium in the process of genome reduction as opposed to** ***S. proteamaculan*****s, which displays genetic hereogeneity.** The chronic nature of *S. entomophila* could be the path of the organism to speciation. Hypervirulent isolates should encode accessory virulence factors that promote pathogenicity while chronic isolates would show the reduced presence of these accessory determinants.

ii) **Systems present in** ***S. entomophila*** **limit horizontal gene transfer.**

iii) **Metabolic and gene regulatory clusters found in** ***S. entomophila*** **promote the chronic nature of the bacterium.**

iv) **The chromosomes encode factors that are important for plasmid functionality**- where they contain elements/regions encoding plasmid maintenance of pADAP variants and limit *S. entomophila* to one pADAP variant.

1.11 Objectives

1. To carry out phylogenetic (Chapter 3), phenotypic and genotypic (Chapter 4) assessment of isolates of *S. entomophila* and *S. proteamaculans* to delineate the taxonomic and pathotypic boundaries of these. Assays will be conducted to determine whether the chromosomes of *S. entomophila* are more uniform.

2. To observe the differential expression of virulence. Key-enzyme assays (chitin, protease, and lipase) will be used to determine whether hypervirulent isolates are up-regulated contrary to chronic disease-causing isolates (Chapters 4 and 7).

3. To describe and define the genome of *S. entomophila* and evaluate potential reduction in the genome in addition presence of unique traits associated with a chronic pathogen (Chapter 5).

4. To undertake whole-genome comparisons of *Serratia* isolates. By constructing annotated datasets for bacterial isolates and species using data developed here (Chapter 5) to investigate differences that may allude to possible secondary elements with advantageous traits (Chapter 6).

5. To determine the influence of chromosomal background on plasmid-encoded disease expression. Drawing on previous research, mitomycin C (MitC) induction experiments to explore mechanisms of regulation. Genomes of MitC conducive and MitC non-inducive isolates will be compared to identify differences in regulatory regions of known virulence clusters. Bio infectivity, growth, and production of accessory virulence factors in the ability to confer full pathotype and to investigate the relationship between plasmid and chromosome (Chapter 7). Further transcriptome comparisons will then be undertaken to compare transconjugant plasmids in non-native chromosomes against their wildtype plasmid-free counterparts for differential pADAP gene expression and accessory virulence genes (Chapter 8).

Chapter 2

General Methods

2.1 Bacterial strains and plasmids

Bacterial strains and plasmid vectors used in this study are listed in Table 2. For overnight cultures, a toothpick of a selected colony was inoculated into 3 mL of Luria-Bertani Broth (LB) in McCartney bottles with appropriate antibiotics and incubated at 200 rpm, at either 30^{o}C for *Serratia* spp. or 37^{o}C for *E. coli* strains, in a Ratek orbital mixer incubator. Agar plates (further described in section 2.3.2) were incubated at 30^{o}C for *Serratia* spp. and 37^{o}C for the bacterial host *E. coli* strains respectively. Unless stated henceforth, the techniques described here forth will be used throughout the project for multiple objectives.

Table 2 Host bacterial strains and plasmid vectors used in this study

Name	Genotype	Reference
E. coli strains		
DH10β	*Δ*(*ara-leu*) *7697 araD139 fhuA ΔlacX74 galK16. galE15 e14- ϕ80dlacZΔM15 recA1 relA1. endA1 nupG rpsL* (*StrR*) *rph spoT1 Δ*(*mrr-. hsdRMS-mcrBC*)	Grant et al. (1990)
ST18	*Escherichia coli* S17 λ*pir* Δ*hemA*	Thoma and Schobert (2009)
EC100D pir^{+}	F– mcrA Δ(mrr-hsdRMS-mcrBC) φ80dlacZΔM15 ΔlacX74 recA1 endA1 araD139 Δ(ara, leu)7697 galU galK λ– rpsL nupG pir-116(DHFR)	Metcalf et al. (1994)
Plasmids		
pGEM T-Easy	Cloning vector, LacZ site, Amp	Promega
pUC19	Cloning vector, LacZ site, Amp	Yanisch-Perron et al. (1985)
pACYC184	Cloning vector, Tet, Cm	Chang and Cohen (1978)
pHP45	Cloning vector, Spec	Moore (1995) NCBI
pJP5603	Suicide vector, Tet	Riedel et al. (2013)

Specific antibiotics for isolate growth were added in concentrations listed in Table 3. For the growth of *E. coli* strain ST18 50 µg/mL of 5-aminolevulinic acid was added to supplement. X-gal was used for blue/while selectivity for pGEM and pUC-19 cloning vectors.

Table 3 Antibiotics and growth substrates used in this study

Antibiotic/ substrate	*E. coli* µg/mL	*Serratia* µg/mL
Ampicillin (Amp)	100	100
Tetracycline (Tet)	30	30
Chloramphenicol (Cm)	30	90
Spectinomycin (Spec)	100	100
Kanamycin (Kan)	50	100
X-gal	100	100
Ala[1]	50	-

[1] Only for growth of *E. coli* strain ST18

2.2 Isolates of *Serratia* spp. investigated

Isolates of *Serratia* spp. investigated in this study are shown in Table 4. Previous classifications of pathogenicity, further described in Chapters 3 and 4, were characterised either by Mark Hurst (AgResearch Ltd, New Zealand) [unpublished data] or by Steve Dodd in the PhD thesis 'Horizontal transfer of plasmid-borne insecticidal toxin genes of *Serratia* species' (Dodd 2003). Further isolates were also obtained from species in the AgResearch bacterial library (Table 4).

The selection of isolates for sequencing (shown in Table 4) was based on differing geographic origin, variations in disease presentation, and a subset of plasmid-free isolates. In addition, the *S. entomophila* isolate A1M02, previously used in the biological control product Invade® was also included.

Table 4 Isolates of *Serratia* used in this study.
Isolates were obtained from the AgResearch Culture Collection (Lincoln New Zealand). Isolates differing in geographic origin and pathotype, as well as isolation date, are further listed in Appendix B.1

Species	Isolate
Serratia entomophila	158[2], 176[2], 192, 204, 207, 210[2], 219[1], 220[2,3],294[2], 295, 305[2,3], 306, 307, 311, 314, 315, 328, 329, 340, 341, 345[1], 359, 364[2,3], 398[2], 399, 400, 401, 438, 440[1], 442[2], 451, 468, 477[2], 480, 482[2,3], 485, 498, 562, 566, 625, 626[1,2], 766, 1021, 1086, 1100[2], 1235, 1343[2,3], 1360, 1373, 1554, 1651, iDIA[2], A1M02[2,3], Sent1[2], Man3[2], I[2], MC2[2]
Serratia proteamaculans	4[2], 142[1,2], 143[1,2], 145[2], 149[1,2], 299[2], 336[2], 465[2], 495, 1048[1], 1071[2], 1129[2], 1137[2], 1457[2], 1769[2], 20093[2], 3041[2], 12a[2], 12newD[2], 25e[2], 28F, 20e, G[2], K[2], Puna18[2], Sprot5, CfB[2], M[2], AGR96X[1,2], LC[2], 10novel[2], MH5[2]
Serratia marcescens	1557[2]
Serratia liquefaciens	F28[2], 377[2], 376[2]
Serratia ficaria	457[2]
Yersinia frederiksenii	Y49[2]

[1] Indicates isolates that were selected for PacBio sequencing
[2] Indicates isolates that were selected for Illumina sequencing
[3] Indicates those sequenced isolates that were selected during this study for their geographical or pathogenic interest or past importance.

2.3 Microbiological methods

2.3.1 Cell growth

Optical density at 600 nm (OD_{600}) was used in parallel with colony-forming units (CFUs), as determined through serial dilution and plating on agar plates, as measures of cell density. For OD's over 1, a 1:10 dilution in 10x phosphate-buffered saline (PBS) was undertaken and absorbance was determined using a Bio-Rad SmartSpec Plus spectrometer. Measurements were undertaken in triplicate with a blank between, and the average was taken.

For the determination of Colony-forming units (CFUs) a serial dilution series in PBS was used. One hundred µL of cell culture was diluted in 900 µL 1x phosphate-buffered saline (PBS) to create a 10^{-1} dilution (Dulbecco and Vogt 1954). This solution was then vortexed before 100 µL was removed with a fresh pipette tip and diluted into a further 900 µL of PBS, to prepare a 10^{-2} dilution. For overnight cultures, serial dilutions of 10^{-6}-10^{-8} were plated on LB agar in duplicate to obtain cell counts. Fifty µL of each dilution was plated onto LB agar and incubated overnight at 30^{o}C.

2.3.2 Plate assays

Unless stated for plate assays bacterial colonies were patched in duplicate and incubated at 30^{o}C. Isolates 626 (*S. entomophila)* and AGR96X (*S. proteamaculans*) were used are standard controls.

2.3.2.1 Selective *Serratia* identification plates

Selective plates were used for the discerning of *Serratia* colonies on media. Primarily, caprylate thallous agar (CTA) was used to identify the presence of *S. entomophila* and *S. proteamaculans* (target species of this study), then itaconate plates were used to differentiate *S. entomophila* from *S. proteamaculans*. DNase plates were also used as a supplement, as S*. entomophila* has a larger halo response size than *S. proteamaculans* (O'Callaghan 1993).

2.3.2.2 Measuring DNase production

Toluidine blue (0.05 g) was dissolved in 100 mL of dd.H_20 and autoclaved, then aseptically mixed with 42 g of Remel DNase test agar (Thermo Scientific) in 900 mL of dd.H_20 to produce 1 L of DNase agar for testing.

2.3.2.3 Measuring Lipase production

Lipase agar was made by adding 10 g peptone, 5 g NaCl, 0.1 g $CaCl_2.2H_2O$, and 20 g Bacto agar with 10 mL Tween 80 (1% w/v) in 1 L dd.H_20. After autoclaving, to avoid denaturing of the emulsifying agent, 10 mL Tween 80 was added and mixed thoroughly before pouring.

2.3.2.4 Measuring Protease production

Milk agar plates were used to determine the ability of an isolate to secrete functional proteases that cleave casein. Milk agar plates were made from two separate components, to avoid denaturing the milk-based protein. Alternate autoclaved parameters of 10 min at 121^{o}C were used following the protocol described by Son (2014) to prevent caramelisation of the sugars. Component one: 10 g milk powder in 250 mL dd.H_20 was autoclaved at a reduced temperature of 116^{o}C for 15 min. Component two: 7.5 g Bacto agar made up to 250 mL with dd.H_20 and autoclaved as previously described. Post autoclaving, the two components were mixed in a 1:1 ratio immediately before pouring.

2.3.2.5 Assessing Cell Motility

Isolates were inoculated onto motility plates and compared to other isolates to determine comparative motility scores. Motility plates were made as described by Son (2014). Agar was made from 5 g tryptone, 2.5 g yeast extract, 2.5 g NaCl, 1.5 g Bacto agar (0.3% w/v) and dd.H_20 up to 500 mL. After autoclaving 40 mL plates were poured and used fresh to avoid loss of moisture.

2.3.2.6 Measuring Chitinase production

Colloidal chitin suspension was prepared from Sigma practical grade crab shells for the preparation of chitin detection assay medium according to the methods described by Rodriguez-Kabana et al. (1983). Ten g of crustacean chitin flakes were ground using a Retsch laboratory mill at 0.5 mm fineness for 30 s and then added to 150 mL of concentrated 12 M HCl. The solution was then left to stir overnight at 20^{o}C. The dissolved solution was then added to a 2 L container containing 1 L of dd.H_20. A suspension was then allowed to form before more water was added to a volume of 1.7 L, from where a precipitate of intermediate chitin substrate was collected on the surface of the supernatant, which was slowly poured off. Tap water was then added to wash and re-suspend the chitin substrate, and the process was repeated eight times using dd.H_20, over four days. pH was then assessed to determine whether the acid had been sufficiently washed off (target pH of 5.5-6.0). The final suspension was then filtered through Whatman Grade 1 filter paper to remove larger particles. The resulting suspension was then stored as a paste in the dark at 4^{o}C.

Chitinase agar was prepared using 500 mL of Bacto agar with minimal salts, a medium adapted from the Hsu and Lockwood (1975) method (Na_2HPO_4 6 g/L; KH_2PO_4; 3 g/L; NH_4Cl; 1 g/L NaCl; 0.5 g/L yeast 0.5 g/L) with 2% chitin added. The solution was then autoclaved and stirred vigorously before plates were quickly poured into standard Petri dishes.

2.3.2.7 Measuring siderophore production

For siderophore production, a plate assay was used to measure iron-chelating by utilising Chromazurol S dye by *Serratia* bacterium. Strong iron-chelating would correspond to a colour change. Plate assays were used as established by (Schwyn and Neilands 1987). Production of CAS plate involves a three-step solution. Solution one: 50 mL dd.H_20 with 60.5 mg of Chromazurol S and 10 mL of iron chloride ($FeCl_2$) solution in 10 mM HCl. Solution two contained 72.9 mg hexadecyl-trimethyl-ammonium bromide dissolved in 40 mL of dd.H_20. Finally, solution three with 6.4 g Na_2PO_4; 0.3 g KH_2PO_4; 2.5 g NH_4Cl; 30.24 g of piperazine-N, N'-bis (2-ethanesulfonic acid) (PIPES), 15g of Bacto afar dissolved in 850 mL dd.H_20. The pH of solution three was then adjusted to 6.8 with NaOH before autoclaving. Twenty mL 20% glucose and 30 mL 10% casamino acid was then added to solution three. Solution two was then added to solution one before the combined solution was added to solution three. Agar was then poured, with colour changing to blue when set.

2.3.2.8 Measuring Phosphatase Production

Pikovskayas agar (PKV) was used for the detection of phosphatases that could be beneficial for plant growth promotion. One L of media was made by adding 10 g glucose; 0.1 g $(NH_4)2SO_4$; 0.2 g KCl. 5 g $MgCl_2.6H_20$; 0.25 g $MgSO_4.7H_2O$, 5 g Tricalcium phosphate ($Ca_3(PO_4)_2$) or hydroxyapatite with 15 g Bacto agar before dissolving in 800 mL of dd.H_20. The pH was then adjusted to pH 7 before making up the volume to 1 L using dd.H_20 and autoclaving before pouring (Pikovskaya 1948).

2.3.3 Indole-3-acetic acid expression

Indole-3-acetic acid (IAA) assays were performed to assess the expression of plant beneficial IAA. IAA detection assays involve the detection of auxins through the use of Salkowski's reagent- a solution of 2% 0.5 M $FeCl_3$ and 35% perchloric acid ($HClO_4$) (Donati et al. 2013). Cultures were grown in tryptic soy broth (Merck) overnight before inoculation of 500 mL into 10 mL of growth medium (Appendix A.1). The inoculated growth medium was then covered and incubated in the dark for 72 h at 200 rpm. Post incubation, samples were pelleted at 6010 × g for 10 min and the supernatant was collected and transferred into a 96 well plate in triplicate. The blank comprised 500 µL of tryptic soy broth plus 10 mL of the growth medium, replicating the test conditions. Sixty µL of Salkowski's reagent was added per well and incubated in the dark for 30 min. A Spectrostar nano microplate reader was used to detect colorimetric activity at 560 nm absorbance in triplicate and the average taken.

The next step was to create a standard curve based on IAA solution recorded at different concentrations. IAA (Sigma) was dissolved in EtOH to form a 1 mg/mL stock solution. Dilutions of concentrated IAA in dd.H_20 were assayed using the above IAA absorption assay and plotted to determine the concentration of IAA µg/mL. The calibration curve was used to determine the value of

'Y' in the line of best fit, which can then be used to determine the concentration for each isolate when absorbance at 530 nm is divided by 'Y'.

2.3.4 Megaplasmid Visualization

Megaplasmids were visualized using a method that uses high temperatures that degrades the chromosomal DNA, leaving behind covalently stable plasmid DNA (Kado and Liu 1981). Overnight cultures were grown in 3 mL of LB broth in McCartney bottles, at 30^{o}C 250 rpm. An aliquot of 300 µL was then taken and spun down for 2 min at 15871 × g. The supernatant was aspirated, and the cell pellet was then resuspended in 2× E buffer (Appendix A.2). Two hundred µL of lysis buffer was then added before the tubes were gently inverted to mix and heated at 55^{o}C for 60-90 min. After incubation, the DNA was then extracted using 300 µL of phenol-chloroform, before inverting the microcentrifuge tube x50 and spinning at 15871 × g for 15 min. Fifty µL of clear supernatant was loaded with 5 µL loading dye per isolate onto a 1% agarose gel in Tris-acetate-EDTA (TAE) buffer. The gel was then run for 240 min at 80 volts, before post-staining in ethidium bromide (EtBr). The gel was then visualized using an EtBr-specific filter on Uvitec Unidoc visualizer. *S. entomophila* isolate A1MO2 with the plasmid pADAP (153 Kb) was used as a standard as well as *S. proteamaculans* isolate AGR96X, which has a similar, but smaller plasmid (120 Kb).

2.3.5 Curing of plasmids

Heat curing was undertaken to remove the plasmid pADAP from isolates of *S. entomophila* to allow for the insertion of pADAP-type plasmids from *S. proteamaculans* to assess the chromosomal effects of alternate plasmids on their virulence. Overnight cultures of target isolates were incubated at 30^{o}C. Dilutions in LB broth at 10^{-5}, 10^{-6}, and 10^{-7} alongside a full-strength sample were patched onto LB agar plates and incubated for 5 d at 42^{o}C. If growth was observed these colonies were patched and assessed via plasmid visualisation (described in section 2.3.4). If no growth was observed, repeats were undertaken, and the final incubation time lowered by 0.5^{o}C each cycle until growth was observed. If plates were overgrown, the inverse was done, and the temperature was increased by 0.5^{o}C until cell growth was around ten cells per plate.

In addition to heat curing, SDS curing was also used as a method to try to cure the plasmid of pAfpX and pADAP baring isolates (Buckner et al. 2018). Overnight cultures of isolates with plasmid tags were supplemented with 10% SDS and incubated at elevated temperatures of 37^{o}C. Serial dilutions were plated on LB agar from the cultures at 16, 24, and 48 h. Colonies grown were then re-patched in tandem on LB agar and LB with appropriate antibiotics for the plasmid tag. Nil growth on the plasmid-specific plate but growth on the LB agar plate would indicate the cassette, and theoretically, the

plasmid is no longer present. Plasmid visualisation gel would then be performed to confirm any plasmid loss.

2.3.6 Growth curves

Growth curves for plasmid conjugants were calculated over 24 h to determine whether the presence of an introduced plasmid compared to wildtype had any overall effect on the growth or metabolism of an isolate. Wildtype isolates with native plasmids and chromosomal transconjugants were grown in 3 mL of appropriate broth overnight. Greiner CELLSTAR® 96 well plates were used to allow for replicates and dilutions for all conjugates and wildtype controls alongside appropriate broth to be used as a blank control. Each isolate was observed in duplicate up to 10^9 dilutions in a single plate assay over 24 h. Spectrostar Omega was used to take readings at 560 nm every 15 min, with constant shaking at 200 rpm. Output data was retrieved from MARS data analysis software and provided in excel format, with raw and blank corrected data.

For itaconate mutant assessment, standard growth curves were undertaken in LB and M9 minimal salts medium (Appendix A.4). Initial overnight cultures were grown for approximately 16 h. OD_{600} was taken and starting concentrations were equalised to a similar CFU. A 500 µL of culture was then inoculated into three flasks of LB broth per isolate, and three flasks of M9 per isolate. Initial time point 0 dilutions of the overnight culture were taken to establish starting concertation, plating onto LB agar. Subsequent timing points where 1 mL was removed per culture flask were at 1 h, 2, 4, 8, 16, 24, 26, and 48 h at each timepoint, OD_{600} was taken in triplicate. Duplicate count plates were spread at two dilutions per flask with dilutions based on the results of OD_{600}.

2.3.7 Microscopy

Fluorescent microscopy was used to determine any effect that conjugation between cells and exotic plasmids had on characteristics of the bacterial cells relative to size, shape, and quantity over an incubation period of approximately 16 h. Observations were undertaken using an Olympus BX50 light microscope for standard light and fluorescent microscopy. Three µL of cell culture was pipetted onto a slide and observed under phase/contrast conditions.

2.3.8 Mitomycin C Induction protocol

Freshly streaked plates of selected isolates on LB media were incubated at 30°C. After ~16 h, 3 mL overnight cultures of each strain with appropriate antibiotics were grown at 250 rpm 30°C. One mL of the overnight culture was used to inoculate a 150 mL flask. The cultures were then grown to an undiluted OD_{600} of 1.1 which took approximately 2.75 h. Cells were then harvested at 2268 × g for 10 min at room temperature. Pellets were resuspended in 0.4 mL LB, adding appropriate antibiotics, and

50 µL of stock Mitomycin C (2 mg/mL). At 6 h post-induction, a booster of 25 µL Mitomycin C was added per flask. The flasks were then incubated overnight at 40 rpm at room temperature. Lytic threads should be observable approximately 4 h post-induction. To prepare the lysate, in a fume hood 0.5 g $NaCl_2$ was added to each flask and 600 µL chloroform and the solution vigorously vortexed for 20 s. The sample was left to stand for three min, and then vortexing was repeated for three bursts. After the final vortex, the solution was left to stand for 5 m. The remaining supernatant was then decanted into a centrifuge tube leaving the chloroform behind. Before ultracentrifugation 3 µL DNase and RNase (Roche, 20 mg/mL) were added and the microcentrifuge tube was inverted and incubated at 37^{o}C for 15 min with occasional inversion. These were then ultra-centrifuged at 18,000 × g for 30 min to remove contaminating flagella and cell debris and the supernatant transferred to a sterile microcentrifuge tube.

As a further analysis for MitC induction of virulence factors, selected isolates underwent assessment of the filtered supernatant. One mL aliquots of supernatant were collected from isolates A1M02 and 626. To each supernatant, tet_{30} was added and the sample filter sterilised using a 45 µm filter before the suspension was used in larval challenges (Chapter 7, section 7.2.1).

2.4 Molecular protocols

2.4.1 Genomic DNA sample preparation

Selected isolates were grown overnight at 30^{o}C with 250 rpm shaking in 3 mL of LB broth for approximately 16 h. Of these overnight cultures, one mL was collected and spun at 6010 × g for two min; and the pellet washed in sterile milli-Q water. The spin and wash steps were then repeated. Genomic DNA or plasmid DNA was extracted using the Bioline ISOLATE II Genomic DNA kit and the High Pure Plasmid Isolation Kit (Roche) following the manufacturers' guidelines respectively. The quality of prepared genomic DNA was assessed by agarose gel electrophoresis (1%) at 95 v for 45 m in TAE buffer. NanoDrop 1000 (Thermo Fisher Scientific, USA) spectrophotometry to determine the concentration (ng/µL) and purity (260/280 nm). PCR amplicon purification was undertaken using the High Pure PCR Purification Kit (Roche, Germany) as per the manufacturer's instructions.

2.4.2 RNA sample preparation

Before sampling, initial time points ranging from 8-20 h were taken to determine optimal harvest time for 2.0-5.0 x 10^9 CFU's for RNAseq. This was done to match the CFU for each comparable isolate so accurate comparisons of sequencing results could be undertaken.

Overnight cultures were prepared with appropriate antibiotics. Flasks were inoculated with 50 µL of overnight culture and aliquots were taken at 14, 16, 18, and 20 h post-induction. Serial dilutions were undertaken and 50 µL of 10^{-6}, 10^{-7}, and 10^{-8} dilutions were plated onto LB to determine CFU. OD_{600} was measured at each time point at a 1:10 dilution with the OD_{600} absorbance reading at <1.

Once the optimal time of 9-10 h was selected, overnight 3 mL cultures of isolates of interest were incubated with appropriate antibiotics at 30^{o}C shaking at 200 rpm. RNAprotect Bacterial Reagent (QIAGEN, Germany) was used following the manufacturer's instructions to obtain the bacterial pelleted sample. RNA isolation was then performed using the RNA mini kit following the manufacturer's instructions. Following an on-column digest, an off-column digest was also performed before the kit clean-up protocol was performed and the RNA isopropanol precipitated overnight. Pellets were air-dried at 37^{o}C for 30 min before being resuspended in RNase-free water and then quantified using nanodrop and Qubit. Six mg/µL concentration of the sample was then added to a centrifuge tube and the liquid evaporated in a SpeedVac. Quality control was performed by Macrogen before sequencing at their facility (South Korea).

2.4.3 Plasmid mini-preparation method

Standard alkaline lysis plasmid mini-preps were conducted using overnight cultures in LB broth of the colony of interest with respective antibiotics at 30-37^{o}C and 200 rpm. One mL of cultured broth (if *E. coli*, 0.5 mL for *Serratia*) was then centrifuged at 30 s 15,871 × g. The pellet was then resuspended in 150 µL of solution one (Appendix A.3.1) before 200 µL of solution two was added (Appendix A.3.2) and mixed. This was then incubated at 37^{o}C for 4 m before 150 µL of NaAc (pH5.2) was added. The tubes were mixed and incubated on ice for 5 m. The sample was then spun at 15,871 × g for 10 min before the supernatant was aspirated, and the pellet dried in a 37^{o}C incubator for 30 min. The final dried pellet was then resuspended in 30 µL of sterile Milli-Q water and stored at -20^{o}C.

2.4.4 BOX A1R PCR

BOX- fingerprinting, first identified by Martin et al. (1992), was undertaken following the standard protocol described in Isaeva et al. (2010). Individual isolates were streaked onto LB agar one day before BOX-PCR and incubated at 37°C overnight. Template DNA was then prepped for each isolate by diluting a toothpick of a bacterial colony into 500 µL of distilled water. This was then vortexed for approximately 10 s to ensure efficient mixing of the diluent.

PCR reagent was prepped using 12.5 µL Reddy mix per reaction in addition to 9 µL of water and 2.5 µL 10 µM BOX A1R primer (Dombek et al. 2000). Reagents were then spun for approximately 5 s to ensure mixing of the solution. Twenty-four µL per bacterial isolate were aliquoted into PCR tubes and then

spiked with 1 µL template DNA for each sample. The BOX-PCR was then run using BIORAD thermocycler (Bio-Rad Laboratories, Inc., CA, USA) using the following cycle parameters: 94^{o}C for 12 m, 35 cycles of 94^{o}C for 1 m, annealing temperature was 56^{o}C for 30 s, and 72^{o}C for 1.5 m, with a last cycle of 72^{o}C for 7 m. PCR products were run on a 1.2% agarose gel stained with 0.7 µL Red safe for 50 m at 95v before being visualised using the Uvitec Unidoc Geldoc Imager (Uvitec, UK).

2.4.5 Standard PCR protocol

Standard PCR mix was followed as shown in Table 5. Platinum Taq DNA polymerase (Invitrogen, USA) was used following the manufacturer's instructions in a 50 µL reagent mix volume.

Table 5 Standard PCR protocol reagent mix for a 50 µL reaction

Reagent	Volume (µL)
10× PCR buffer	5
$MgCl_2$	2
dNTP mix (10 mM)	1
Primer (f) 100 µM	0.5
Primer (r) 100 µM	0.5
Template DNA	1
Platinum Taq Polymerase	0.2
Nuclease-free water	38.8

Primers used in this study are listed in Table 6.

Table 6 Primers used in this study
Sequence restriction sites are shown underlined where present.

Name	Sequence (5'-3')	Description	Template	Restriction site
tetNcoI_f	aaaccatggGAGTTAGTCTTGAAGTCATGCGC	Tet cassette, Ita	pACYC184	NcoI
tetNcoI_r	aaaccatggGCATTCACAGTTCTCCGCAAG	Tet cassette, Ita	pACYC184	NcoI
Ita_f	aaatctagaGGTTTGTACCCGCCGTTAGCA	Ita region	626	XbaI
Ita_r	aaatctagaCTCGCCCTTGACGGCCTGATCG	Ita region	626	XbaI
repA_f	TGGAGGGGAACGATCTTCTTGAGG	Universal repA		
repA_r	GCCCCACTTTCTTGTACCATCCAG	Universal repA		
BOX A1R	CTACGGCAAGGCGACGCTGACG	BOX fingerprinting		

Ita_gene_F	gatatcCAGATCATCGAATCCCACCGT	Gene complement		EcoRV
Ita_gene_R	gatatcGTGGTTGGCGCATCTCCC	Gene complement		EcoRV
Ita_operon_f	gatatcGTGGTTGGCGCATCTCCC	Operon complement		EcoRV
Ita_operon_r	gatatcACGCTAAAAACGGCCCCC	Operon complement		EcoRV
Itaconate_val_f	CGCGCTGATTGAAAGCGCA	Ita validation	626	
Itaconate_val_r	GGCCCGATGTTTCACGAAA	Ita validation	626	
BacF	AGGTGACGGTGGTAATGG	Region A		
BacR	TGGAAGTGCATATCCATA	Region A		
Din_1	GCGCTGCTGACCGACTTCTC	Use with Din_3	626	
Din_2	aaagagctcCCGCCAAACAAAGCGACGAG	Use with Din_5, final	626	SacI
Din_3	cggttttctggaaggcgagcaGGTTTACCTGTAATGCTGATTG	Use with Din_1	626	
Din_4	tcaagccgacgccgcttcgcgCGGGAACATACTGCCTTCGC	Use with Din_6	626	
Din_5	aaagagctcCCTTCATCTATCACCTGGTCAGAGCTCTTT	Use with Din_2, final	626	SacI
Din_6	GGCAGTTCAGGCACTGCTTC	Use with Din_4	626	
Din_val_f	GCATTCACAGTTCTCCGCAAG	Validation- use with Din_5	626	
Din_val_r	ACATCGACCCACGGCGTAAC	Validation- use with Din_2	626	
pHP45Omega_Spec_amp_F	AAACCCTCACTGATCCGCATG	Spec cassette	pHP45	
pHP45Omega_Spec_amp_R	AGACTTGACCTGATAGTTTGGCTG	Spec cassette	pHP45	
M13_F	GTAAAACGACGGCCAGT	Sequencing	pGEM	
M13_R	GCGGATAACAATTTCACACAGG	Sequencing	pGEM	
M13_pUC_r	GTTTTCCCAGTCACGAC	Sequencing	pUC19	
M13_pUC_r	CAGGAAACAGCTATGAC	Sequencing	pUC19	
27_f	AGAGTTTGATCMTGGCTCAG	16S sequencing		
1492_r	TACGGYTACCTTGTTACGACTT	16S sequencing		
AfpX_f	GGGCAGTATTCCGCTGCAGACGG	Validation of pAfpX insert		
AfpX_r	TCTAGCCATCATGGTGCGGCAACC	Validation of pAfpX insert		

Underscore denotes restriction enzyme binding site. Lowercase letters aaa-tail.
F indicates forward primer and R indicates reverse primer.

2.4.6 Colony PCR

Pooled colony samples in 500 µL of dd.H_20 were used to screen large quantities of genome recombinants. Single colonies were patched to a fresh agar plate and then into a microcentrifuge tube.

Pooled samples were vortexed and 1 µL was used as template DNA using DreamTaq (Thermoscience, USA) following the manufacturer's instructions. PCR cycling was then optimised following the DreamTaq protocol.

2.4.7 Fusion PCR

To mutate regions encoding *dinI* further discussed in Chapter 5, section 5.2.8, fusion PCR was used as a method to incorporate an antibiotic cassette in the PCR amplicon to replace the gene of interest (Szewczyk et al. 2006). Fusion PCR utilizes an 'up' and 'down' stream PCR fragment to insert an antibiotic cassette into a region of interest where selective restriction enzyme sites are not required for mutagenesis (Figure 5).

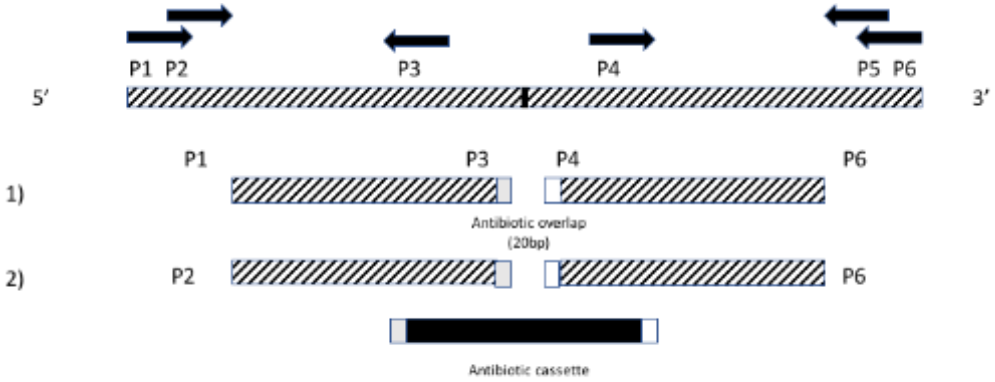


Figure 5 Three step fusion PCR diagram shown
P3 and P4 should have a 20 bp overlap to antibiotic cassette selected 1) Initial amplification of upstream and downstream sequences from gene of interest of approximately 1Kb size 2) PCR antibiotic cassette of choice out of plasmid vector 3) using primers 2 and 5 use all three fragments to construct a fusion PCR product. Primers 2 and 5 should have a poly-A tail at both the 5' and 3' end to aid in transformation protocol. Image modified from (Szewczyk, Nayak et al. (2006).

Once the fusion PCR was successful, as determined by gel visualization of fragment sizes and sensitivity to a selected antibiotic cassette, the insert was ligated and transformed into the PCR cloning vector pGEM-T Easy following the protocol in sections 2.4.9 and 2.4.10. Plasmid DNA was then subjected to restriction enzyme (RE) digestion. Gel electrophoresis of the digest (section 2.4.12) was then undertaken, and where a fragment of the expected size was found, the fragment was sequenced by Sanger sequencing at Macrogen to confirm the insert. The cloned insert was then ligated into an appropriate suicide vector. Due to the site of the clones used in this study, vector pJP5603 was used for cloning protocols. The cloned insert in the suicide vector was then electroporated into EC100D following protocol in section 2.4.11. Purified DNA was then electroporated into *E. coli* ST18 and the transformed bacterium selectively grown on LB plates + antibiotics + ALA. Verification of the insert size by RE digestion was then undertaken before conjugation into the appropriate *Serratia* strain following

the protocol in section 2.4.15. Colonies from successful conjugation were then patched onto selective plates, with validation PCR used to confirm the insert of the cassette into the isolate.

2.4.8 Restriction enzyme-based mutagenesis

For some regions, fusion PCR protocol proved difficult, either due to the primer sets or the presence of AT-rich genomic regions. For these regions, a standard mutagenesis approach using specific restriction enzyme sites positioned in the gene of interest. By identifying two restriction sites in the targeted gene, primers were made for an appropriate antibiotic cassette with flanking restriction enzyme sites to allow for the insertion of a cassette to disrupt gene function. Resultant ligation and transformants were undertaken as outlined in sections 2.4.9 and 2.4.10. This method was utilised to mutate the itaconate operon in isolate 626 with a tetracycline cassette.

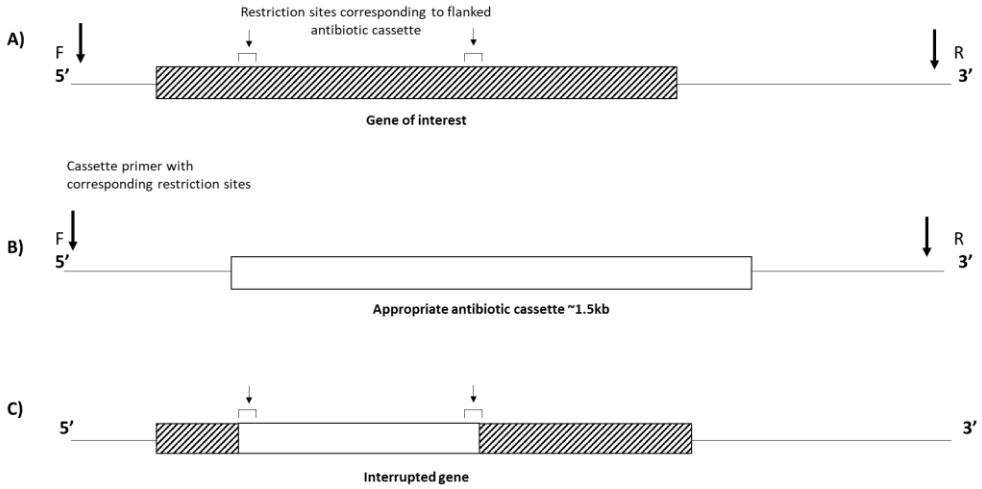


Figure 6 Schematic of restriction site targeted mutagenesis
A) Target region amplified flanking the gene of interest. Restriction sites identified on the target gene that correspond with antibiotic cassette. B) Antibiotic cassette amplified using primers from Table 6. C) Restriction digest and ligation of gene of interest and antibiotic cassette result in interrupted gene with relevant antibiotic marker.

2.4.9 Ligation

Cloning into a plasmid vector of interest involved the following steps. For the fusion PCR method, the region was PCR amplified, with 3.5 µL of purified PCR product ligated into pGEM (5 µL 2× ligation buffer, 1 µL Promega T4 DNA ligase, 0.5 µL pGEM T-easy DNA). This was then left at room temperature (~22^{o}C) overnight before transformation. Other ligations were performed using Invitrogen ligase with the following protocol. For a 10 µL reaction, 2 µL 5× ligation buffer, vector DNA (1-2 µL) and the purified digested mini-prep or genomic DNA (3-5 µL) at a respective 1:3 ratio were added, and mix increased to 9 µL with sterile dd.H_20 before adding 1 µL Invitrogen T4 DNA ligase and left overnight.

2.4.10 Chemical transformation

Ligations were transformed into chemically competent DH10b cells. Three µL of the ligation was chilled on ice for 5 min before 100 µL of competent cells were added. This was then heat shocked at 42^{o}C for 30 s before being returned to ice for 30 s. One mL of LB broth was then added, and the microcentrifuge tube was placed horizontally at 37^{o}C for 1 h. The microcentrifuge tube was then centrifuged at 6010 × g for 3 m before the supernatant was aspirated, leaving approximately 100 µL to resuspend the pellet. The resuspended pellet was then divided and spread over two LB plates (80 µL and 20 µL respectively) containing the appropriate antibiotics and the plates incubated overnight at 37^{o}C. For pGEM, plates were selected with X-gal for blue or white colonies and Ampicillin.

2.4.11 Electro transformation

For transformation into suicide vector, electro-competent EC100 cells were used. Cells were prepared and electroporated using the method of Dower et al. (1988). One hundred µL of electrocompetent cells were placed on ice before 1 µL of chilled DNA was added and transferred to an electroporation cuvette Electroporation was carried out at 2.5 kV with 200 Ω resistance. One mL of LB broth was added to the cuvette before being transferred to a microcentrifuge tube and incubated at the required temperature (30-37^{o}C) for 1 h. Post incubation, the mix was then spun for 6010 × g for 3 m, before aspiration of supernatant leaving 100 µL to resuspend the pellet. The resuspended pellet was divided and spread over two LB plates (80 µL and 20 µL respectively) containing the appropriate antibiotics and the plates incubated overnight at the required temperature.

2.4.12 Restriction digestion for mutation

To insert a DNA fragment into the vector of choice, the DNA from the vector and insert were digested with an appropriate restriction enzyme at the required temperature. For the digest, 2-4 µL of plasmid DNA or 20 µL of PCR product was used, with 10 µL of Cut Smart buffer (NEB, USA) added. The solution was made up to 99 µL with dd.$H_2$0 before 1 µL of the required restriction enzyme added and the reaction incubated at 37^{o}C for 1.5 h.

2.4.13 Ethanol precipitation

To precipitate the DNA, following restriction enzyme digestion, 2 µL of pellet paint and 10 µL or 3 M NaAc was added to the 100 µL of digest reaction. Two hundred µL of ethanol was then added and the solution briefly vortexed. This suspension was then incubated at room temperature (~22^{o}C) for 2 m before centrifuging at 6010 × g for 5 m. The supernatant was aspirated, and the resultant pellet dried for 20 min at 37^{o}C. The pellet was resuspended in 7 µL of sterile dd.$H_2$0. Two µL of the digest was loaded onto 1% agarose gel (buffer) with a 10 Kb ladder to determine the quality of the digest and the quantities required for ligation.

To determine the successful insertion of a vector and insert, a diagnostic digest was performed. One µL of Cut Smart buffer was added to 3 µL of DNA (or 1 µL plasmid DNA) with 0.3 µL of restriction enzyme. This was incubated at 37^{o}C for 1.5 h before being loaded onto a 1% agarose gel, electrophoresed for 45 min at 95 V with TAE buffer against a 10 Kb ladder to determine by their size whether the insert and vector had been inserted.

2.4.14 Chromosomal and plasmid tags

pJP5603 suicide vectors of tetracycline and spectinomycin antibiotic cassette constructs in 'region A', an intergenic region not thought to alter function, were previously constructed and validated by Connor Watson (AgResearch and Bio-Protection Research Centre, Lincoln University) and used in this study to conjugate into the chromosomes of interest. Plasmid backbone tags were designed and created by Lesley Sitter (AgResearch and Bio-Protection Research Centre, Lincoln University), and provided in strains of A1M02, 626, and AGR96X. Chloramphenicol was used as a backbone tag and used as a positive marker for conjugation of pADAP into various chromosomal backgrounds.

2.4.15 Conjugation

Conjugation was used to insert the chromosomal tag into isolates of interest, as well as the pADAP and pAfpX plasmids into tagged chromosomal backgrounds. To undertake conjugation 3 mL overnight cultures of recipient and donor strains were inoculated with appropriate antibiotics (plus ALA in the case of *E. coli* ST18) and incubated overnight. One mL aliquots of each were then centrifuged at 6010 × g for 3 m. The supernatant was aspirated and then the pellet was resuspended in 1 mL of LB broth. For ST18 1.5 µL of ALA was added. The suspensions were then pelleted at 6010 × g for 5 min before aspirating and resuspending in 500 µL of LB broth. For ST18, 0.75 µL of ALA was then added. One hundred µL of each (recipient first) was then pipetted onto an LB agar plate. The plate was then swirled briefly before being incubated upright at 30^{o}C for 6 h. Microcentrifuge tubes of washed cultures for both donor and recipient were also incubated. After 6 h, 1 mL of broth was added to the plate before a hockey stick spreader was used to inoculate three plates with appropriate antibiotics for the transconjugant. For the controls, 50 µL of the donor and recipient were independently spread onto plates to select for the transconjugant. If control colonies grew the plates were discarded, and the method was repeated.

Conjugates were validated by PCR- where primers based in the flanking conserved region identified on the chromosome where the spectinomycin and tetracycline cassettes were to be inserted. If present, there was an expected increase of approximately 1.5 Kb in the amplification product. Once PCR amplification showed a potential tagged strain, primers were designed for overlapping the antibiotic

cassette to amplify the 'up' and 'down' regions to be sequenced. Sequenced reads were then used to show whether the corresponding antibiotic cassette had successfully inserted into the correct region. pADAP conjugation was further verified by undertaking *repA* specific PCR amplification using repA primers (Table 6) and observing banding against wildtype pathogenic and non-pathogenic isolates.

2.4.16 Validation and stability of plasmid transconjugants

Plasmid visualisation methods previously described were utilised to validate the plasmid transconjugant. Once these validation tools showed the correct plasmid insert and chromosomal BOX profile, the stability of the transconjugant was assessed by detection of growth on the plasmid and chromosomal specific antibiotic plates at time points of 24, 48, and 72 h after inoculation into a flask containing LB broth with no antibiotic selection. In these experiments three plates for each conjugate were used, being chromosome-specific plate, plasmid specific, and combined chromosome + plasmid specific selection plates. Plate counts were then used to determine the CFU/mL of cells with the selective tag. Growth on all three to a similar CFU/mL would determine that the plasmid was stable in the chromosome at all three time points, whereas growth on all but the combined selective plate would show that plasmid loss had occurred.

2.4.17 Sequence validation

DNA sequencing was performed by Macrogen (South Korea) using the Standard-Seq sequencing service for sequence validation of plasmid vector inserts. Plasmid vectors were prepared as per the manufacturer's instructions using the High Pure Plasmid Isolation Kit (Roche). Reads were sequenced using the M13 primer set for pUC19 to validate the insertion of PCR amplified regions. DNA sequences were aligned to the corresponding genome using the 'map to reference' function of Geneious Prime 2019.2.1 software, a software used for the analysis of sequencing data (Apweiler et al. 2004, Eddy 2011).

2.5 Bioinformatic analysis

Fifty-six sequences of *Serratia* spp. and one *Yersinia* isolate were used in this study across Chapters 5 and 6. One isolate of *S. marcescens*, three of *S. liquefaciens,* and an *S. ficaria* were included for comparative genomics. Twenty-three isolates of *S. entomophila* and 28 *S. proteamaculans* were also sequenced as part of this study.

2.5.1 PacBio sequencing assembly

Raw PacBio data was received from Macrogen in three separate read files per strain, and in three separate formats: Bam5, FASTQ, and FASTA. The three raw PacBio FASTA reads were initially merged and then assembled using Canu to formulate complete genomic contigs (Koren et al. 2017). Although Canu has a built-in error correction, corrections using Illumina short reads needed to be undertaken to increase the accuracy of small base changes. This was done using Pilon (Walker et al. 2014). Output files for assembled contigs and the corrected reads were inputted into Circlator, software utilised for circulation of genomic contigs to correct any overhangs as a feature of PacBio read assemblies circular consensus reads (Hunt et al. 2015). Chromosome orientation was based on the *dnaA* replication gene.

2.5.2 Illumina sequencing assembly

Illumina sequencing was undertaken by Macrogen (South Korea). The quality of raw data was assessed using FastQC (Andrews 2010) and trimmed using Trimmomatic (Bolger et al. 2014). Illumina sequences were assembled using A5-miseq (Coil et al. 2015). Genomes were then visualized using IGV and scaffolded using SSPACE before being corrected by Gapfiller (Boetzer et al. 2011, Boetzer and Pirovano 2012). The list of sequenced assembled isolates is available in Chapters 3, 5, and 6 (Table 13, Table 24, and Table 25).

2.5.3 Removal of plasmid sequences from genomes

For PacBio sequenced isolates, the Circlator output separated the genome into chromosome and plasmid-specific contigs, and therefore the removal of plasmids was simple and involved the deletion of the extra scaffolds from the fasta file. After processing, these identified plasmids were verified using nBLAST, and correlated with the presence/ absence data from plasmid visualization data of *S. entomophila* and *S. proteamaculans* isolates alongside work previously completed by Dodd (2003) and Sitter (2020).

For Illumina sequences, pADAP sequences from PacBio reads were added as a custom BLAST in Geneious Prime. Plasmids to remove were identified by analyzing previous plasmid visualizations undertaken by Hurst [unpublished data] and in this study and identified plasmid lists through Sitter (2020) and Dodd (2013). The contigs could then be blasted (Basic Local Alignment Search Tool, BLAST, NCBI, USA) for the identity of reads to the plasmid, and those showing high identity (~90% sequence identity) and a corresponding size could be removed. Typically, a single contig contained the plasmid.

2.5.4 Annotation of assembled sequences

PROKKA (Seemann 2014) was used to annotate the corrected and trimmed reads to be used for comparative genome analysis. Corrected and assembled DNA sequences in FASTA format were used

as an input with multiple feature prediction tools utilised to formulate an accurate coding gene annotation. PROKKA utilises the BLAST+ core database and HMM libraries to perform full annotations from UniProt bacterial protein lists which can then be manually corrected in Geneious Prime. Prodigal predicts coding sequence coordinates (Hyatt et al. 2010). RNAmmer, Aragorn, and Infernal were utilised for the identification of various RNA coordinates (Laslett and Canback 2004, Lagesen et al. 2007, Kolbe and Eddy 2011). SignalP was used to identify signal leader peptides (Petersen et al. 2011). If no match was found for coding sequences, the gene was labelled 'hypothetical protein.'

Assembled and annotated reads were imported into Geneious Prime where full annotations were visualized, and individual regions of interest explored. Primers for regions of interest identified were then designed using this program, alongside the identification of regions and 16S rRNA (see section 2.5.15) by the internal custom BLAST function.

2.5.5 Average nucleotide Identity

Average nucleotide identity (ANI) can be used to show how genetically related species of similar bacteria are. In this instance, ANI of PacBio sequenced chromosomes was calculated using Mypathogen ANI tools against reference sequences of other *Serratia* spp. (Han et al. 2016). Percentage ANI below 90% can be described as dissimilar enough to not be of the same species.

2.5.6 BRIG alignment

Genome alignment can also be observed using BLAST Ring Image Generator (BRIG), software used to display BLAST ring images (Alikhan et al. 2011). BRIG is a java-based application utilizing the BLASTn to compare and align genome nucleotides and display them in concentric comparable rings aligned to a central reference genome. Additional features such as G+C skew and content can be added to indicate genomic islands and regions of difference. Using BRIG, aligned genomes can display percentage similarity based on colour scheme- where a solid block colour indicates over 90% percentage similarity between the query and reference, and a lighter shade over 70%. White areas within an alignment show non-homologous regions. Chromosomal contigs of *S. entomophila* and *S. proteamaculans* were aligned comparatively to the chromosome of NCBI reference reads of the respective species to demonstrate the genome conservation across the chromosome in both strains, and to show areas of conservation and divergence. Broad regions of differences within a species were identified. A G+C skew ring was added to display anomalies in G+C content to possibly identify genomic islands.

2.5.7 Mauve

ProgressiveMauve aligner was utilised from the Mauve software to align the chromosomal contigs of PacBio reads via anchored alignments (Darling et al. 2004). ProgressiveMauve can align large numbers

of sequences with high sensitivity for partially conserved regions, utilizing Hidden Markov Model (HMM) to remove erroneous alignments. By selecting the full alignment option, ProgressiveMauve used MUSCLE (MUltiple Sequence Comparison by Log-Expectation) to perform a recursive anchor search which in turn results in a full-gapped alignment (Darling et al. 2010).

2.5.8 Roary

Roary (Page et al. 2015), a pan-genomic pipeline, was used to visualize differences across the genomes in relation to the phylogeny of the sequences. Roary takes annotated GFF files from PROKKA annotations and constructs a presence-absence core and accessory gene list. These were then assembled alongside a Newick tree to form a presence-absence schematic showing areas of conservation and differentiation across the whole genomes. All genomes sequenced here of *S. entomophila* were input to show similarities and differences. Compared to the Mauve alignment tool, this technique is easier to interpret in discerning differences, as the more genomes added to a Mauve alignment the more complex the differences and inversions become to interpret.

The output presence and absence .csv file from ROARY was then run through a processing R script to obtain core genome statistics for all isolates and a presence/absence heat map was generated. From the Roary plots, exported image regions of homogeneity and differentiation were identified between isolates of interest. To determine the gene content of these regions, the query pan-genome function of Roary was used.

2.5.9 Pangenome analysis

Pangenome bounds were calculated using the R package micropan (Snipen and Liland 2015). Chromosomes of isolates of the same species were first compared, detecting orthologous protein groups using Proteinortho6, a standalone Linux tool (Lechner et al. 2011). A .csv output table delineating the presence and absence of identified protein clusters was used as input for the micropan package. Micropan functions were then used to establish pangenome summary and rarefaction curves, before calculating the bounds of the pangenome with Heap's law and Chao statistic for the estimated number of genes in the pangenome.

2.5.10 Scoary

Complementary to Roary, Scoary- the microbial pan-GWAS (Genome-wide association study)- was used to calculate gene association between genes in the core genome and identified traits (Brynildsrud et al. 2016). Scoary uses a python-based script to score phenotypic traits in a pan-genome. A binary traits file was used to select isolates that shared location, pathogenicity, species, and beneficial factors. These were compiled into a .csv file where the isolate name was matched with the presence and

absence file generated from Roary previously. Permutations were selected based on the number of isolates included in the analysis. Cut-offs used naïve P-value to determine statistical significance as to whether the null hypothesis of no relation to the trait could be deduced.

2.5.11 *Breseq*

For a comprehensive evaluation of genome comparisons, the Linux tool *breseq* was used to analyse all sequenced genomes (Deatherage and Barrick 2014). *Breseq* allows for the identification of additions and deletions between all genomes on a larger scale, giving a comprehensive comparison that can allow for rapid differentiation which can be correlated to species, location, and virulence. To assess differences between closely related strains and different pathotypes- isolates 626 (pathogenic) and 477 (non-pathogenic) were selected due to the similarity of their chromosomal layout and structure.

2.5.12 Islandviewer4

Islandviewer 4 was used to calculate and visualize potential genomic islands (GI's) in the chromosomes of sequenced isolates (Bertelli et al. 2017). Islandviewer4 is an integrated web tool utilises Islandpick, SIGI-HMM, Islander, and IslandPath-DIMOB prediction software to accurately predict the location of GI's on GenBank uploaded chromosomes. IslandPath-DIMOB predictions are based on mobility genes and nucleotide bias of the sequence (Hsiao et al. 2003), whereas SIGI-HMM uses codon bias and Hidden Markov Models (Waack et al. 2006). Finally, IslandPick applies comparative genomics methods (Langille et al. 2008) Fully annotated GenBank files were uploaded to Islandviewer4 server for identification of these regions.

2.5.13 Genome mining

Once all large-scale comparisons were complete, further information was derived from sequenced annotated datasets by identifying genes of known secondary virulence determinants on the chromosomal contigs of each sequenced isolate by *in silico* computational methodologies as outlined below. Additional screening was undertaken on individual genes to determine the function of hypotheticals by inputting amino acid protein sequence into Phyre2 (Kelley et al. 2015).

2.5.13.1 Protease identification

Sequences were screened using the MEROPs peptidase database for known proteases and proteolytic enzymes that are located on the chromosome of each isolate (Rawlings et al. 2010). An index search of the MEROPs database of related *Serratia* organisms was undertaken whereby 187 *Serratia* isolates were found. An isolate of *S. marcescens* and *S. proteamaculans* with complete sequences and a broad list of identified peptidases were selected. Sequence lists of these known peptidases were then downloaded for each isolate. Once these datasets were acquired, NCBI BLAST databases for each

chromosome to be assessed were created in the command line, before tBLASTn query for each FASTA file of peptidase sequences was undertaken against each BLAST database. Output files were then imported into excel. Each file lists a protein identifier coding to a peptidase or homolog listed on the MEROPS database (Rawlings et al. 2010). From this, individual proteases could be identified.

Once imported into excel the putative peptidases were then sorted based on percentage identity; those with less than 60 % similarity to the reference were discarded. These were then compiled into a presence or absence .csv file of all chromosomes (Appendix D.6). The presence or absence.csv file was then manipulated to display unique hits per isolate that could then be identified by their MEROPs protein identifier and manually searched in the online database.

2.5.13.2 Lipase, chitinase, and DNAse identification

The hmmer3 hmmsearch (hidden markov models) function was utilised for the chromosomal identification of motifs indicating the presence of accessory virulence determinants such as lipases, chitinases, and DNases (Eddy 2011). Target Pfam families of HMM domains were identified before HMM profiles were built using the Hmmer3 hmmbuild feature. Families 18 and 19 glycoside hydrolases were searched for to identity chitinases, while Pfam (protein family database) class 1 and class 3 lipases were used for lipase and phospholipase motif predictions. Hmmsearch was then utilised using the faa protein file for each sequence to search for significantly similar matched sequences. Hmmsearch output then ranks the best scoring domains based on an E cut-off threshold. Proteins implicated by the domain HMM search were then manually inputted into Uniprot to confirm the identity of the protein. Hypothetical proteins were further scrutinised by assessment through Phyre2 (Kelley et al. 2015). Manual genome searches were also undertaken on the nucleotide sequence in Geneious prime to identify the genes.

2.5.13.3 AntiSMASH

AntiSMASH is a compiled web-based server that utilises open source tools such as HMMer 3, NCBI BLAST+, PySVG, and FASTTREE to identify and annotate potential secondary metabolite clusters on a bacterial genome *in silico* (Blin et al. 2019). Sequenced chromosomes, with plasmid contigs removed, were fed into antiSMASH using default parameters. 'Relaxed' strictness was used to predict for partially complete biosynthetic clusters. These were compiled into a presence or absence .csv file for all chromosomes.

2.5.14 COGs

COG analysis on isolate 626 was undertaken by Eric Altermann (AgResearch, New Zealand). Five reference sequences from GenBank were also assessed for accurate comparison to other *Serratia* spp.

The results of these were then used to aid in the construction of a genome Atlas for *S. entomophila* isolate 626 that utilised BLASTp and COG annotations to create an accurate display of chromosomal features.

2.5.15 16S ribosomal DNA (rDNA) phylogenies

Phylogenies based on the 16S rDNA gene were constructed from sequences extracted from completed Illumina and PacBio reads, alongside the identification of further isolates of interest for specific 16S sequencing, including 16S sequencing of assembled genomes from section 2.5.4, where nucleotide ambiguities were identified. Further isolates for sequencing sent to Macrogen are listed in Table 7. These were then imported into the MEGA 7 software and aligned via CLUSTAL algorithm (Thompson et al. 2002, Kumar et al. 2016). Approximately 30 bp was trimmed from each end to account for variability in sequence length extracted. Maximum-likelihood trees were then constructed using Tamura-Nei method with 500 bootstraps.

Table 7 List of isolates identified for specific 16S rRNA region sequencing in addition to the original genome sequence

* Denotes isolates that were previously sequenced, and sequencing was repeated to increase the reliability of the 16S gene

Species	Isolate
Serratia entomophila	176, 219, 220*, 314,325, 398, 400, 477*, 482, 626*, 625, F2, NewF, F

2.5.16 Core genome phylogenies

16S rRNA sequencing for closely related species is not the most accurate for depicting relationships and differences due to the short length of the sequence (approximately 1460 bp post trimming). Therefore, a larger range of genes would ideally be selected. Using core genomes consisting of multiple genes created a more accurate view of the relationship between closely related strains and the possible evolutionary route some of these strains are taking. These were initially generated by using the output of a core conserved genome alignment from Roary of the isolated chromosome, with pADAP removed. To then increase accuracy and consider recombination events, these were then repeated using Snippy (Seeman 2015). Snippy identifies single nucleotide polymorphisms (SNPs) between a reference genome and sequenced FASTA files. These were then aligned by SNPs into a full core alignment which was used to build a high-resolution phylogeny using core sites in Gubbins (Genealogies Unbiased By recomBinations In Nucleotide Sequences) (Croucher et al. 2014). The Gubbins command line-based software was then used to formulate a phylogenetic maximum likelihood tree based on these alignments. Gubbins uses FastTree and RAxML to create a more

accurate phylogeny accounting for recombination, mutation, and horizontal gene transfer across similar isolates.

2.5.17 RNA transcriptome analysis

RNASeq transcriptome analysis was undertaken by Charles Hefer (AgResearch, New Zealand).

Initial quality checking was undertaken using FASTQC (Andrews 2010) before being trimmed with Trimmomatic (Bolger et al. 2014). The .gff files generated from the Roary annotation of *Serratia* isolates were then used, with the short read libraries aligned to the reference genome using HISAT2, using default parameters. StringTie (Pertea et al. 2015)was used to assemble novel transcripts, with the transcript count calculated using BALLGOWN (Frazee et al. 2015). Further analysis was undertaken using a COG annotated isolate of *S. entomophila* to profile the change in COG transcriptome of the chromosome and the plasmid when inserted into an unfamiliar background.

2.6 Assaying virulence

2.6.1 Grass grub collection

Third instar grass grub larvae of *Costelytra giveni* were collected from AgResearch, Lincoln, and were kept at 4°C in individual wells of ice cube trays until used. Pre-bioassays were undertaken with larvae added to interval columns in a 60 well ice cube tray. A 3 mm^3 cube of carrot was placed in each well. Trays were then stacked with an empty tray on top. To maintain ambient moisture levels, a saturated paper towel was placed on the top tray, and the trays were then sealed in a plastic bag. Trays were then stored at 15°C for 3 days. Cessation of feeding during the prefeed period would disqualify larvae from use in the following bioassays. Those not feeding or exhibiting disease symptoms were discarded.

2.6.2 Bioassay

Bioassay protocols were based on methods described by Jackson et al. (2001). Selected healthy larvae were fed cubes of carrot (3-4 mm^3 in size) that were rolled on lawn plates of bacterial isolates of interest grown overnight on LB agar at 30^{o}C, with inoculation rates at approximately 10^7 CFU/mL per larvae, per treatment. Each treatment comprised 12 larvae and was undertaken in duplicate. The treated carrot was administered on day one, and fresh tray changes with fresh untreated carrot cubes were given on days three and six. Treatments were run in parallel with negative control grubs fed carrot and the positive controls of the pathogenic isolates A1MO2 and AGR96X. Symptoms of disease (non-feeding, amber discolouration) were visually assessed on days three, six, and 12 with a final day 12 larval observation used to determine disease. For classic amber disease presentation, the most telling sign was amber discolouration, clearance of the gut, and cessation of feeding. Some pathotypes, such as AGR96X hypervirulent isolate present differently, therefore in these cases less colour change

was noted. Due to incomplete paralysis of the mandibles cessation of feeding included small amounts of 'nibbling' around the outer edges of the carrot in hypervirulent disease. These were listed as diseased states for the corresponding isolates.

2.6.3 Dose-response bioassay

Lethal concentration was determined using the above bioassay method layout for each isolate (including transconjugants), with different doses derived from a serial dilution of an O/N culture (~5-9 × 10^{-9} CFU/mL). Lethal concentration (LC_{50}) and lethal time (LT_{50}) were assessed against controls of wildtype non-pathogenic and donor plasmid carrying positive controls against healthy negative controls. Serial dilutions of 10^{-1}-10^{-4} were used, as any lower dilution would not be expected to show any pathogenicity (based on previous preliminary work). Serial dilutions were conducted up to 10^{-8} and then 50 µL plated onto LB to allow for calculation of the initial CFU dose. Twelve larval treatments were undertaken in triplicate and rotated within the ice cube trays to negate any 'edge effect' and any outliers. Ten µL of each cell concentration was applied to a 3-4 mm^3 square of carrot and a single grass grub added. The larvae were checked at daily time points for 12 days for mortality rates and cessation of feeding.

LC_{50} and LT_{50} were calculated using Probit analysis (where **α** is the intercept, and β is the slope)

$$P = \alpha + \beta[log_{10}(Dose)]$$

For LC_{50} final day, 12 readings for each dilution were converted to the proportion of disease to healthy and corrected to the blank prior to analysis. For assessment to be effective the background disease level must be low enough to not interfere with the pathotype readings. Bioassays were undertaken in triplicate and the average of each result taken and presented with the standard error of the mean, where σ is the standard deviation and *n* represents the number of samples ($SE = \frac{\sigma}{\sqrt{n}}$). For LT_{50} the proportion of disease present at each time point was used, before following the same protocol for Probit analysis.

2.6.4 Enumeration of bacteria from larvae

Enumerations were performed to determine bio infectivity and establishment of host colonies in the diseased larval host (Jackson et al. 1993). Larvae were surface sterilised with ethanol (100%) for 1 min, before being rinsed with dd.H_2O. The individual larvae were then dried and weighed, from where they were independently placed in microcentrifuge tubes. To each microcentrifuge tube, the volume was made up to 1 mL with dd.H_20, and a sterile micro pestle was used to homogenise the grub. Serial dilutions were then taken and plated onto appropriate antibiotics or CTA for identification of *Serratia*.

2.6.5 Pathotype determination of bioassays

For bioassays of wildtype isolates, day 12 readings of diseased, dead, and healthy grubs were used. The standard error of the sum of each bioassay was used to calculate disease, whereby pathogenic was defined as <25% of the control 'blank' grubs plus >75% of test grubs were diseased. Hypervirulence was described as >75% of grubs dead at day 12. The 3rd category (mixed) was used where an isolate didn't fit either of the previous pathotypes but still had a combined total of >75% affected larvae.

2.6.6 Protozoa lysate protocol

Four healthy larvae were selected, and the fermentation sacs (rear of the insect hindgut) were removed following the protocol of Zhang and Jackson (2008). The fermentation sacks were then cut open, and 100 µL of PBS was added and mixed by pipetting up and down. The resulting suspension was then added to 500 µL of PBS in a microcentrifuge tube, and a 1:20 ratio layer of mineral oil was added to maintain semi anaerobic conditions (Lam et al. 2018). Five µL of an overnight bacterial culture of *S. entomophila* (~10^8-10^9 dilution) was then added, resulting in a 200 fold dilution of approximately 1×10^6 cells These samples were then incubated at 15^{o}C for 6 h.

Post incubation, bacteria were removed from the protozoa using a gentamycin 400 wash for 1 h (Rønn et al. 2017), and the microcentrifuge tubes were subsequently centrifuged at 500 × g for 4 min. The supernatant was discarded before the pellet was resuspended in 1 mL of PBS and the wash/spin process repeated three additional times. The final resuspension used 1 mL of PBS buffer containing 300 µg mL^{-1} of gentamycin, and the sample was then incubated at room temperature for 1 h. A serial dilution was undertaken to determine the proportion of cells remaining in the protozoa solution. The protozoa were then lysed to determine the presence of internalised bacteria. The solution was spun at 500 × g for 4 min, aspirated and the resultant pellet resuspended in 1 mL PBS and the process repeated three times. The final resuspension used 1 mL PBS, with 0.4% Triton-X to lyse the protozoa. Immediately following the addition of Triton-X the serial dilutions were performed in PBS, and 50 µL of each was plated onto LB agar plates. (2x plates of each dilution). Plates were incubated overnight at 30^{o}C. Colonies grown from both pre lyse and post lyse were then patched onto itaconate, adonitol, and DNase plates to determine whether *S. entomophila* bacteria was present. If any growth on the control plates from pre-lyse were identified as *S. entomophila*, the post lyse plates were also discarded as the gentamycin wash was deemed 'insufficient' to kill non-encapsulated bacteria. This experiment was undertaken in triplicate.

2.7 Statistical analysis

2.7.1 Statistical analysis of the plate assays

A matrix of pathotype vs isolate geography was compiled alongside plate data from results gathered as per section 2.3.2. Initial grouped analysis of variance ANOVAs based on species and geographic isolation was undertaken for each trait. The grouping was categorised with species (including subspecies), location, and pathotype. Species grouping was based on 16S rRNA phylogeny splitting of subspecies to assess differences in known subspecies of *S. proteamaculans.* Using this method, ANOVA was undertaken using 3-way grouping interactions for each variable (assay). Note that only those isolates with a geographic grouping that was mirrored in the opposite species, (e.g., species of *S. entomophila* from Manawatū-Wanganui when species of *S. proteamaculans* are also from Manawatū-Wanganui) could be assessed in this way. For graphical output grouped interval plots were used to show a 95% confidence interval from the mean for groups between species and pathogenicity. Single groupings for each geographic isolation, species, and pathotype were also used to determine single factorial associations between traits and isolates.

2.7.2 Statistical analysis of bioassay data

P-values were generated using a two-sample T-test for bioassay data based on the instance of disease, death, or combined outcome in relation to the negative control (blank) for each assay. Standard error was calculated as previously described in section 2.6.3. Error bars used in bioassay data and bio infectivity assays were generated as the standard error of the mean.

Chapter 3

Delineating species boundaries of *Serratia* spp.

3.1 Introduction

Prior to initiating comparative assays on isolates of *Serratia* used in this study, 16S rRNA phylogenies were used to establish species boundaries for effective delineation between species (Objective 1). A taxonomic range of type strains for species of *Serratia* from across the genus and a variety of ecological niches were used to represent the genetic diversity. For example, *Serratia aquatilus* has been isolated from drinking water systems, whereas *Serratia myotis* was isolated from bats (García-Fraile et al. 2015, Kämpfer and Glaeser 2016). Using a broad range of *Serratia* spp. would allow for the correct placement of all in-house isolates to a correct taxonomic group based on the traditional approach. Initial analysis using 16S phylogenies allows for primary identification and species delineation of in-house isolates based on prior publication data such as that of Garcia-Fraile et al. (2020) alongside ANI utilised in Chapters 5 and 6, before further refinement utilising various comparative genomic methodologies (Chapter 6).

Previous research has utilised 16S rDNA sequencing methods to categorise *Serratia* isolates as members of specific species. For example, (Mc Carlie et al. 2020) used 16S rDNA phylogeny as a primary method to identify a multidrug-resistant *Serratia* sp. strain. Reclassification of *Serratia* isolates as synonyms of one species is not uncommon. Taxonomic relationship of a *Serratia glossinae* to *Serratia fonticola* found that these isolates shared over 99.6% nucleotide similarity, so the species were synonymised (Kämpfer and Glaeser 2015). Before the current study, *S. entomophila* and *S. proteamaculans* were the two species assumed to be involved in amber disease. Therefore, defining their genetic similarities and differences within the context of the bacterial cohort utilised in this study, was important to correctly identify species groupings to accurately determine any traits or genes attributed to a pathotype or isolate.

3.2 Results

To establish the species boundaries currently understood in literature, the 16S rDNA sequences of all *Serratia* spp. type strain isolates from GenBank were used to generate a maximum-likelihood tree to replicate that found in (García-Fraile et al. 2015), shown in Figure 7. It should be noted that during this study *Serratia verspertilionis* was reclassified as a heterotypic synonym of *Serratia ficaria* (Garcia-Fraile et al. 2020), therefore was left out of this analysis.

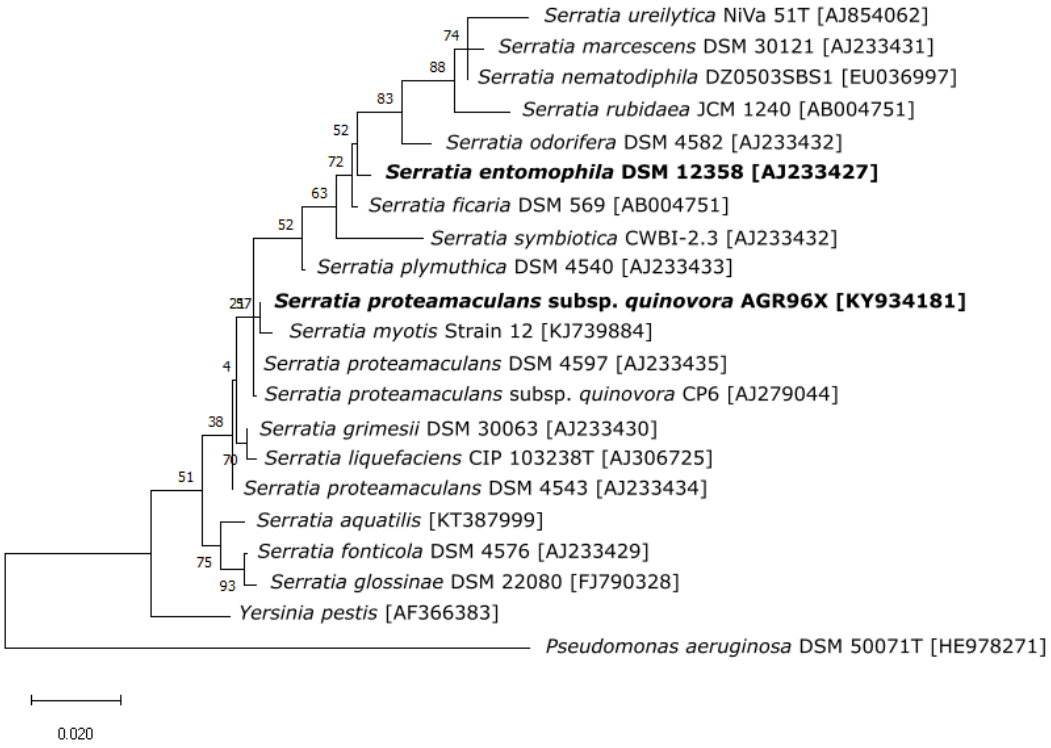


Figure 7 16S rDNA Maximum likelihood tree of sequenced *Serratia* spp. Type strains from GenBank
The evolutionary history was inferred by using the Maximum Likelihood method and Tamura-Nei model in MegaX. The tree with the highest log likelihood (-4038.22) is shown. The percentage of trees in which the associated taxa clustered together is shown next to the branches. The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. This analysis involved 21 nucleotide sequences. Type strains for *Serratia proteamaculans* AGR96X and *Serratia entomophila* shown in bold.

After type strains were utilised to establish the species delineation in the genus, in-house sequencing of isolates was then used to identify the 16S rRNA sequence. With reference to Figure 8, it is evident that the assessed isolates split into clades of both *S. entomophila* and *S. proteamaculans*. Several isolates previously thought to be *S. proteamaculans* were then further identified as *S. entomophila* isolates and reclassified (isolates F2 and Sent1).

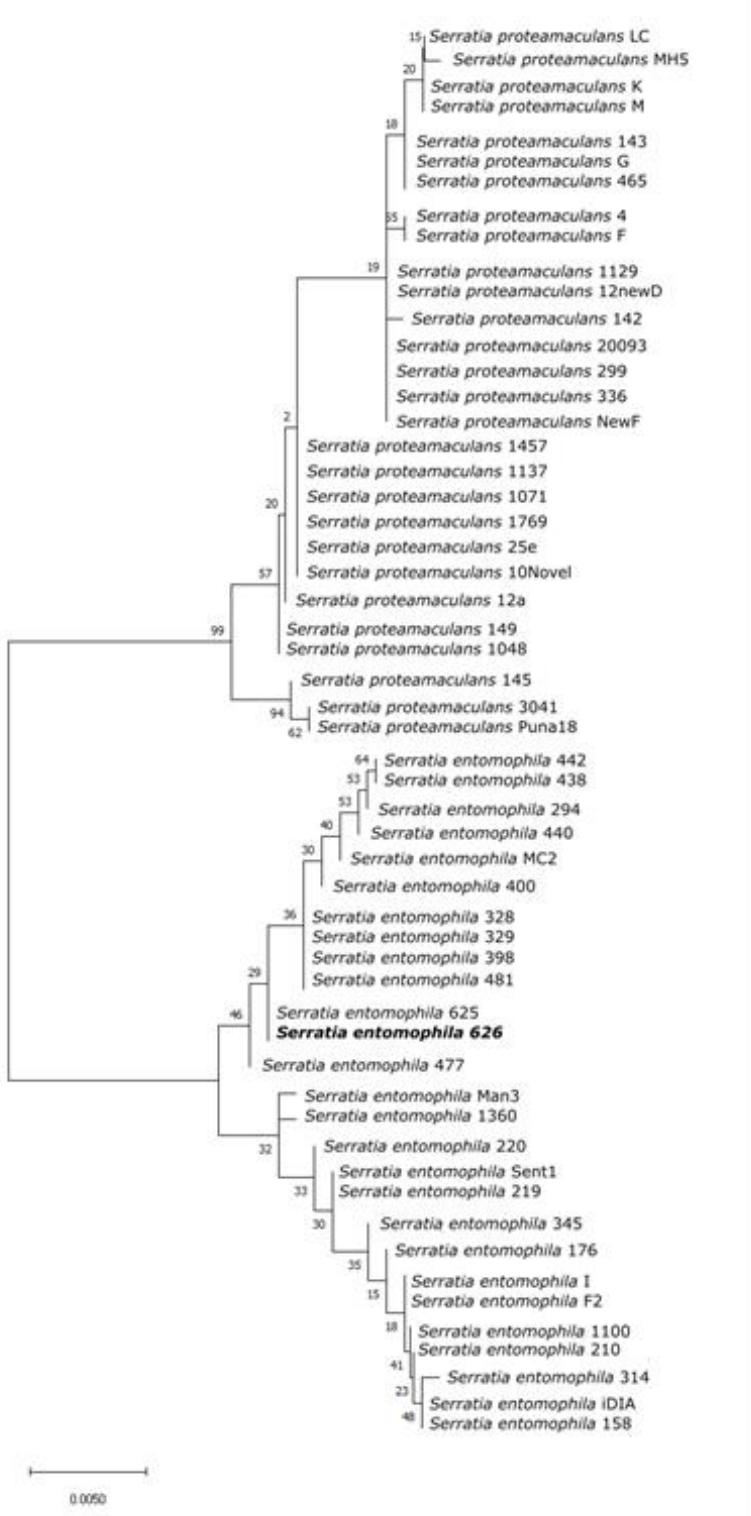


Figure 8 16S rDNA Maximum likelihood tree of sequenced *Serratia entomophila* and *Serratia proteamaculans* in this study

The evolutionary history was inferred by using the Maximum Likelihood method and Tamura-Nei model. The tree with the highest log likelihood (-2287.02) is shown. The percentage of trees in which the associated taxa clustered together is shown next to the branches. Initial tree(s) for the heuristic search were obtained automatically by applying Neighbor-Joining and BioNJ algorithms to a matrix of pairwise distances estimated using the Tamura-Nei model, and then selecting the topology with superior log likelihood value. The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. This analysis involved 55 nucleotide sequences. *Serratia entomophila* isolate 626 is shown in bold.

Sequences of 16S from *S. entomophila* strains isolated overseas were downloaded from GenBank (strain and accession numbers are shown in Figure 9). Together with sequences of other *Serratia* spp. acquired from GenBank (Figure 7) a maximum likelihood tree was generated. The 16S phylogeny confirms that the 23 isolates of *S. entomophila* in this study, alongside three GenBank accessible *S. entomophila* 16S sequences, are represented in one clade. The fourth GenBank sequence was placed close to *S. marcescens* but was unresolved. This could be a laboratory mistake or misidentification, or another very closely but unidentified species. Similarly, *S. vespertillionis* is placed within the *S. entomophila* clade but is also unresolved. The three international isolates that fit into the clade (Mor.4, AB2, and H343) fit well into the *S. entomophila* groupings defined by the type strain and New Zealand isolates. Therefore, based on constructed phylogenies we can restrict the current knowledge of isolation of *S. entomophila* to New Zealand, France, Mexico, India, and South Korea.

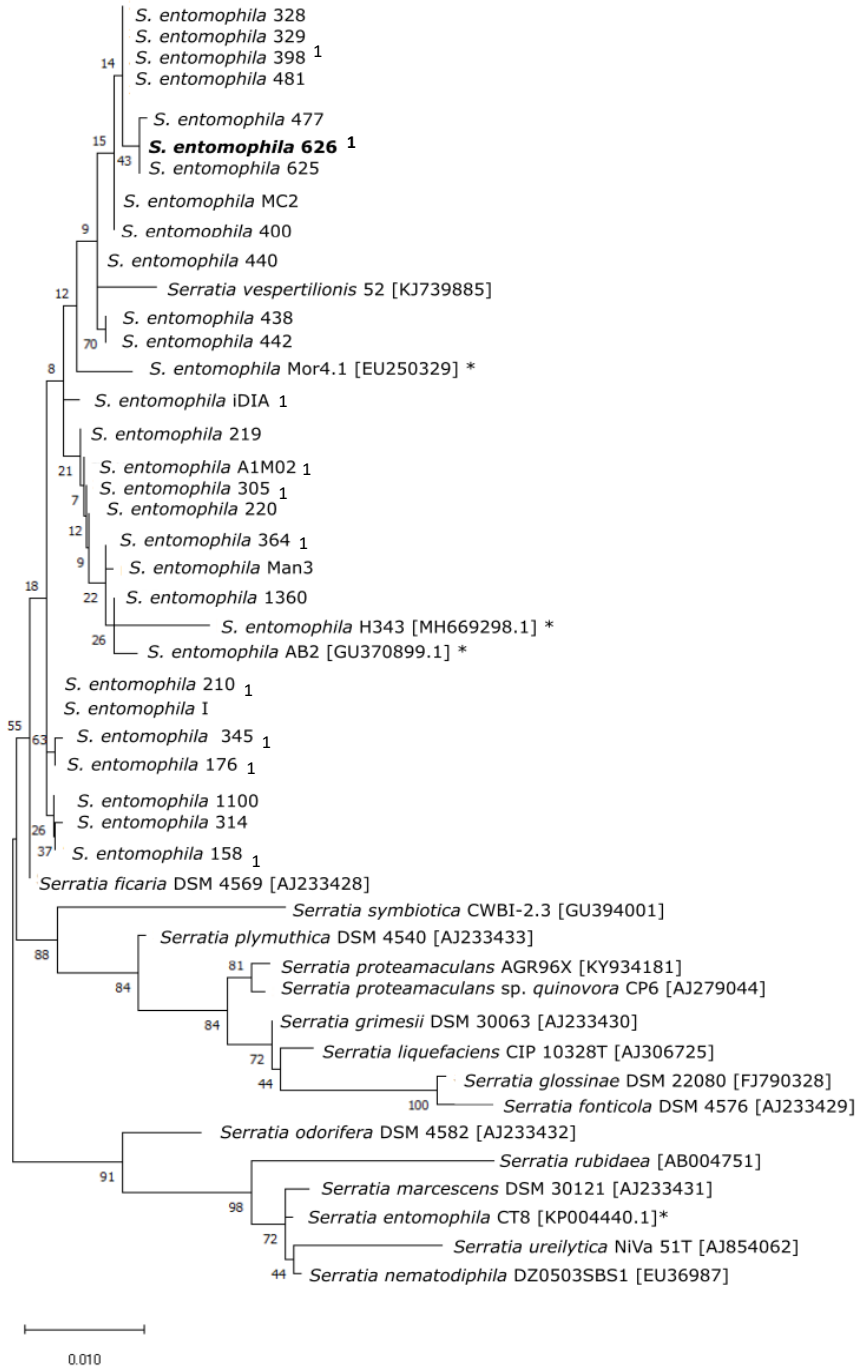


Figure 9 16S rDNA phylogeny for *Serratia entomophila*

The evolutionary history was inferred by using the Maximum Likelihood method based on the Tamura-Nei model. The tree with the highest loglikelihood is shown. The percentage of trees in which the associated taxa cluster is shown near branch nodes. This tree is drawn to scale with branch length measured in the number of substitutions per site. This analysis involved 46 nucleotide sequences and was conducted in MEGAX phylogenetic analysis software. Asterisks denote isolates of *S. entomophila* downloaded from GenBank identified from overseas. Bio-shield® isolate 626 shown in bold. GenBank accessed overseas 16S *S. entomophila* isolates indicated with an asterisk. Isolates AB2 and CT8 were identified in India. H343 in South Korea and Mor.4 in Mexico. Accession codes for each are shown in labels.

[1] indicates isolates of *S. entomophila* that were later in this study classified as pathogenic.

A final phylogeny was then constructed of all isolates of *S. entomophila* and *S. proteamaculans* used in this study, together with those species data in the wider *Serratia* genus (Figure 7), using *Pseudomonas aeruginosa* type strain as the outgroup. Three trees were constructed to compare (neighbour joining, maximum likelihood, and maximum parsimony), with the maximum likelihood tree shown in Figure 10 representing the consensus topology. Results from this tree illustrate that within the *Serratia* there is low bootstrap support for separation of the different species, suggesting that 16S alone is unable to confidently resolve them. However, clades are largely consistent with the taxonomy as it is currently accepted. Importantly for this study, *S. entomophila* and *S. proteamaculans* remain distinct from each other. With the inclusion of species across the genus, *S. entomophila* now appears polyphyletic and these groupings may need to be considered later in terms of pathogenicity. For this study and based on these findings it was decided that the isolate of *S. ficaria* would be initially grouped with *S. entomophila* for accurate comparison. In addition, the species of *S. proteamaculans* splits into two defined groups corroborate the literature of *S. proteamaculans* subspecies.

From this tree, species delineations were then generated to be used throughout this thesis resulting in three main comparative groups and defined in more detail in Table 8:

I. *S. entomophila* and *S. ficaria*
II. *S. proteamaculans* subsp. *proteamaculans*
III. *S. proteamaculans* subsp. *quinovora*

Table 8 Isolates defined into each grouping as defined by 16S rRNA phylogenies
Isolates shown are isolates to be further used in *in vitro* and *in silico* comparative analysis

Group	Species	Isolate
1	*Serratia entomophila* & *Serratia ficaria*	158, 176, 204, 207, 210, 305, 307, 340,341, 359, 364, 398, 451, 468, 480, 485, 562, 626, 1100, 1651, iDIA, A1M02, Sent1, 192, 219,220,294, 295, 306, 311, 314, 315, 328, 329, 345, 399, 400, 401, 438, 440, 442, 477, 482,498, 566, 625, 766, 1021, 1086, 1235, 1343, 1360, 1373, 1554, Man3
2	*Serratia proteamaculans* subsp. *quinovora*	12a, 20093, 3041, Puna18, 145, 1048, 149, 1071, 1137, 1769. 25e, 10Novel, AGR96X
3	*Serratia proteamaculans*	336, 299,1129, 12newD, 4, F, 142, 465, G, 143, LC, MH5, K, M

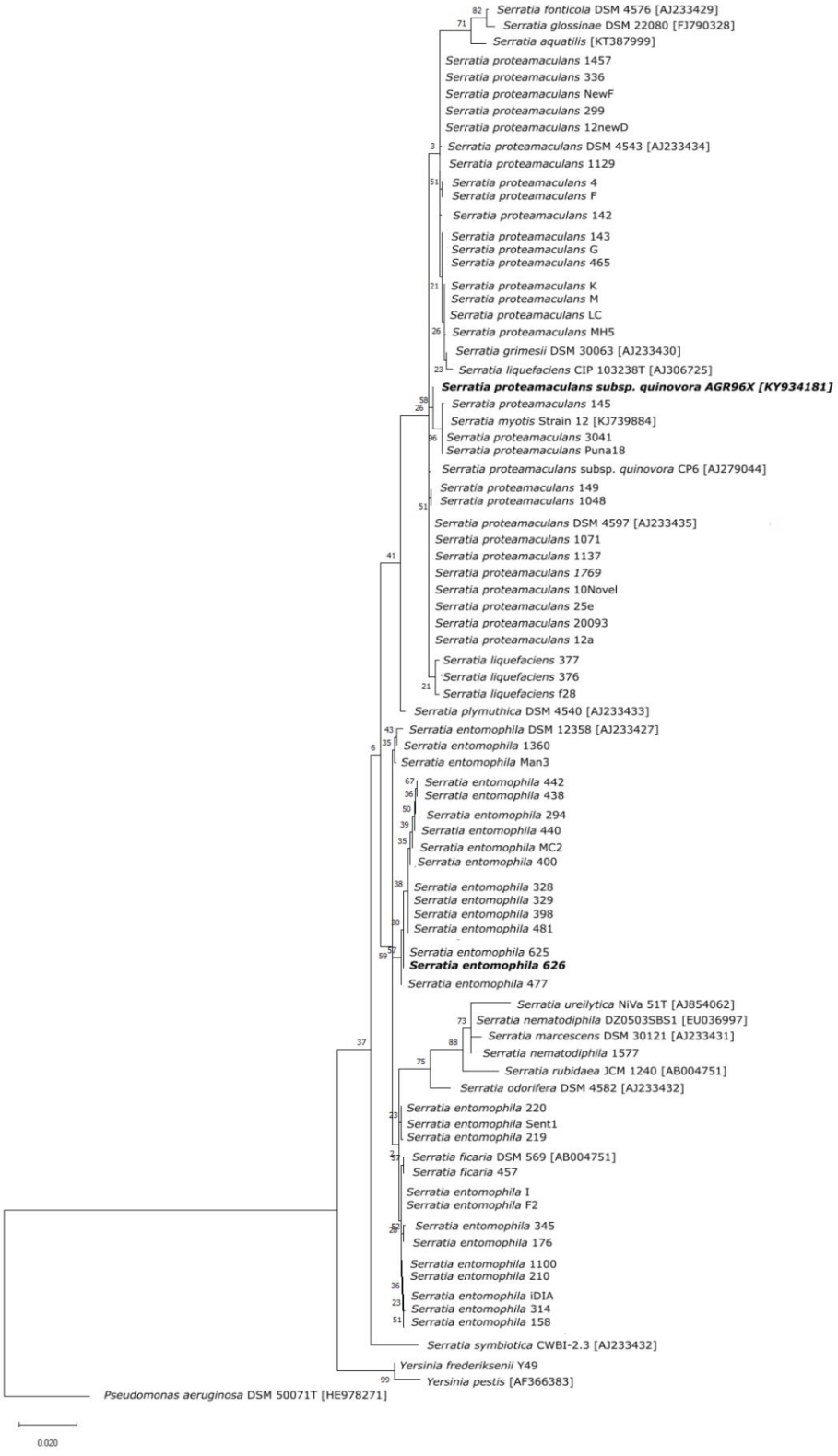


Figure 10 16S rDNA phylogeny for *Serratia* isolates and type strains with final classifications
The evolutionary history was inferred by using the Maximum Likelihood method based on the Tamura-Nei model. The tree with the highest loglikelihood is shown. The percentage of trees which the associated taxa cluster is shown near branch nodes. This tree is drawn to scale with branch length measured in the number of substitutions per site. This analysis involved 82 nucleotide sequences and was conducted in MEGAX phylogenetic analysis software. Asterisks denotes isolates of *Serratia entomophila* downloaded from GenBank identified from overseas. Bio-shield® isolate 626 and *Serratia proteamaculans* hypervirulent isolate AGR96X shown in bold. The tree has been rooted to *Pseudomonas aeruginosa* type strain as an out grouping.

3.3 Discussion

Based on previous literature, isolates collated for this study were described as either *S. proteamaculans* or *S. entomophila*, with initial studies identifying both species as the causal agents of amber disease, with the species further delineated by plate assays (Glare et al. 1993, O'Callaghan and Jackson 1993). Previously conducted 16S rRNA phylogenies by García-Fraile et al. (2015) also indicated that there were two species. However, further differentiation with additional genes is needed to establish the distinct species boundaries, as the polyphyly illustrated within the genus means it is not so clear cut.

These analyses indicate that the species descriptions currently available for the *Serratia* genus may not be fully complete. *Serratia ficaria*, first described by Grimont et al. (1979) and assigned as a novel species as a result of its classification into a new DNA hybridisation group, actually seems to be incorporated into the *S. entomophila* species according to 16S. In the case of *S. proteamaculans*, the subspecies *quinovora* had previously been defined by biochemical characterisation (Grimont et al. 1982), with isolate AGR96X used in this study already designated a taxonomic member of this subspecies by Hurst et al. (2018). Previous research by Grimont et al. (1982) described this species with *S. grimesii* and *S. liquefaciens* as '*S. liquefaciens*-like' to encompass that there was no carbon utilisation approach to differentiate between the three species.

Although 16S is useful for indicating species boundaries in bacterial isolates, limitations have been recognised beyond the suitability of its evolutionary rate for all taxa. Bacterial isolates can harbour more than one ribosomal rRNA operon, therefore the accuracy of 16S specific sequencing and phenotyping depends on the amount of heterogeneity within these bacterial genomes (Acinas et al. 2004). In further work in Chapters 5 and 6, the number of ribosomal rRNA operons diverges greatly between isolates of the *Serratia*. However, Acinas et al. (2004) described sequence divergence in the 16S gene as being less than 1%. In similar organisms such as those described as the same species, single nucleotide differences between operons could allow for phylogenetic misinterpretation. The numerous replicates of the 16S gene present in *Serratia* could allow for variation of encoded nucleotides presenting as differentiation. To expand on this similarity, Fox et al. (1992) showed that 16S analysis in *Bacillus* spp. was not a distinguishing enough criterion to determine species identity, where 99.5% 16S sequence identity did not correspond with similarity defined by DNA-DNA hybridisation. In the case of extremely related isolates, such as those used in this study, that are potentially isolates of a single species, delineation by 16S would then cast doubt on the degree of similarity. MultiLocus Sequence Analysis (MLSA) has been suggested as a complementary analysis to 16S to define closer related isolates. López-Hermoso et al. (2017) found that MLSA correlated the

species definition level defined by DNA-DNA hybridisation with 97% similarity in concatenated MLSA sequences of the *gyrB, recA*, *rpoA* and *rpoD* genes, where 16S could not differentiate into species groups. Utilising whole genome tools such as MLSA will be addressed further in Chapters 5 and 6 to support the species groupings designed above (section 3.2).

Overall, it's clear that based on the literature the cross-over between previously characterised species and new isolates have blurred the current lines of the species within the *Serratia* genus. For this study, the results acquired from the 16S rDNA phylogenies resulted in *S. entomophila* and *S. ficaria* isolate 457 first being analysed as one group. *Serratia proteamaculans* will be divided into two groups, one encompassing the predicted isolates sitting in the clade of *S. proteamaculans* subsp. *quinovora*, alongside identified previously *S. liquefaciens* isolates.

Chapter 4

Bacterial phenotyping and genotyping of the *Serratia* spp.

4.1 Introduction

Phenotypic characteristics can be used to define bacterial production of proteases or extracellular DNases. These can then be correlated with the taxonomic assignment of species and pathogenicity of a bacterium. By identifying patterns on a macro-scale, we can begin to infer relationships of specific traits to evolutionary processes- be it by geographic isolation or beneficial adaptation to an ecological niche. This has been achieved in previous studies where taxa distribution has been linked to biogeographic patterns such as the prevalence of a host species (Hanson et al. 2012). Testing for intra-taxon biogeographic traits is not a new concept and has been utilised previously to determine global phylogeography of the bacterium *Polynucleobacter necessaries* subspecies (Hahn et al. 2015). These prediction processes are utilised globally in investigative studies to determine invasion patterns of pathogenic microorganisms, demonstrated by the database of Invasive Forest Pathogens (IFPs) in Europe, which predicts invasion probabilities based on country-specific land use (Santini et al. 2013). By applying this rationale to the entomopathogens *Serratia entomophila* and *Serratia proteamaculans,* we can potentially determine if common characteristics are associated with geography and land use. This could be influential in the prevalence of pathogenic or avirulent strains in the environment.

Isolates that encode an array of virulence accessory determinants, such as chitinases or proteases, may have increased bacterial pathogenic capacity resulting in an increased infectivity lethality. In human hosts, chitinases have been reported as enhancers of bacterial infectivity, while proteases are thought to cause a heightened inflammatory response, resulting in further damage (Chaudhuri et al. 2010, Marshall et al. 2017). Linking specific pathogen characteristics to particular niche opportunities will support and drive hypotheses on the evolution of these insect pathogens.

In this Chapter, the main phenotypic characteristics of *S. entomophila* and *S proteamaculans* isolates will be defined using plate assays (Chapter 2, section 2.3.2) and bacterial genotyping (Chapter 2, section 2.4.4). The correlation of the presence of the pADAP plasmid with pathotypes assigned using standard bioassays is confirmed (Chapter 2, section 2.6.2).

4.2 Results

4.2.1 *Serratia* spp. isolates

A cohort of 95 isolates of various pathovars of *S. entomophila* and *S. proteamaculans* were sourced from the AgResearch culture collections on-site in Lincoln (Table 4). The majority of these isolates used were previously characterised with restriction enzyme profiling and bioassays to determine pathogenicity by Dodd (2003) or by Hurst (unpublished data). Several isolates, such as the Chatham Island (Rēkohu) isolates 294, 440, and 442 were unique to this study. Although the collection of bacterial isolates from the AgResearch library is extensive, with isolates sourced throughout New Zealand, the majority of isolates were locally sourced typically from field trial sites in the Canterbury region.

Bacterial isolates from diverse geographical locations (including overseas isolates of *S. entomophila* and *S. ficaria* from France) were first examined via genotyping and plate assays to distinguish any relationship between geography and conserved species evolution, relating to pathotype or the expression of secondary virulence factors (Chapter 3). In this study to ensure sample diversity, isolates were selected based on differing geographic locations and pathogenicity. Various isolates of additional *Serratia* spp. were used as a comparison, alongside *Yersinia frederiksenii* isolate Y49 as a closely related non-*Serratia* out-group.

The geographic source of New Zealand isolates used in this study is depicted in Figure 11. It is important to note that these isolates are specific to latitude and longitude of known locations from extensive collections over the past 20 years. As these are now being analyzed in this study on a large scale as one geological dataset, coordinates are specific to a field site while older samples may only be to a district. In addition to New Zealand, there are two isolates from France (*S. entomophila* 220 and *S. ficaria* 457) and one from Argentina (*S. proteamaculans* 495) that are not shown that were also phenotypically assessed and used as outlier samples from the main cohort of New Zealand isolates.

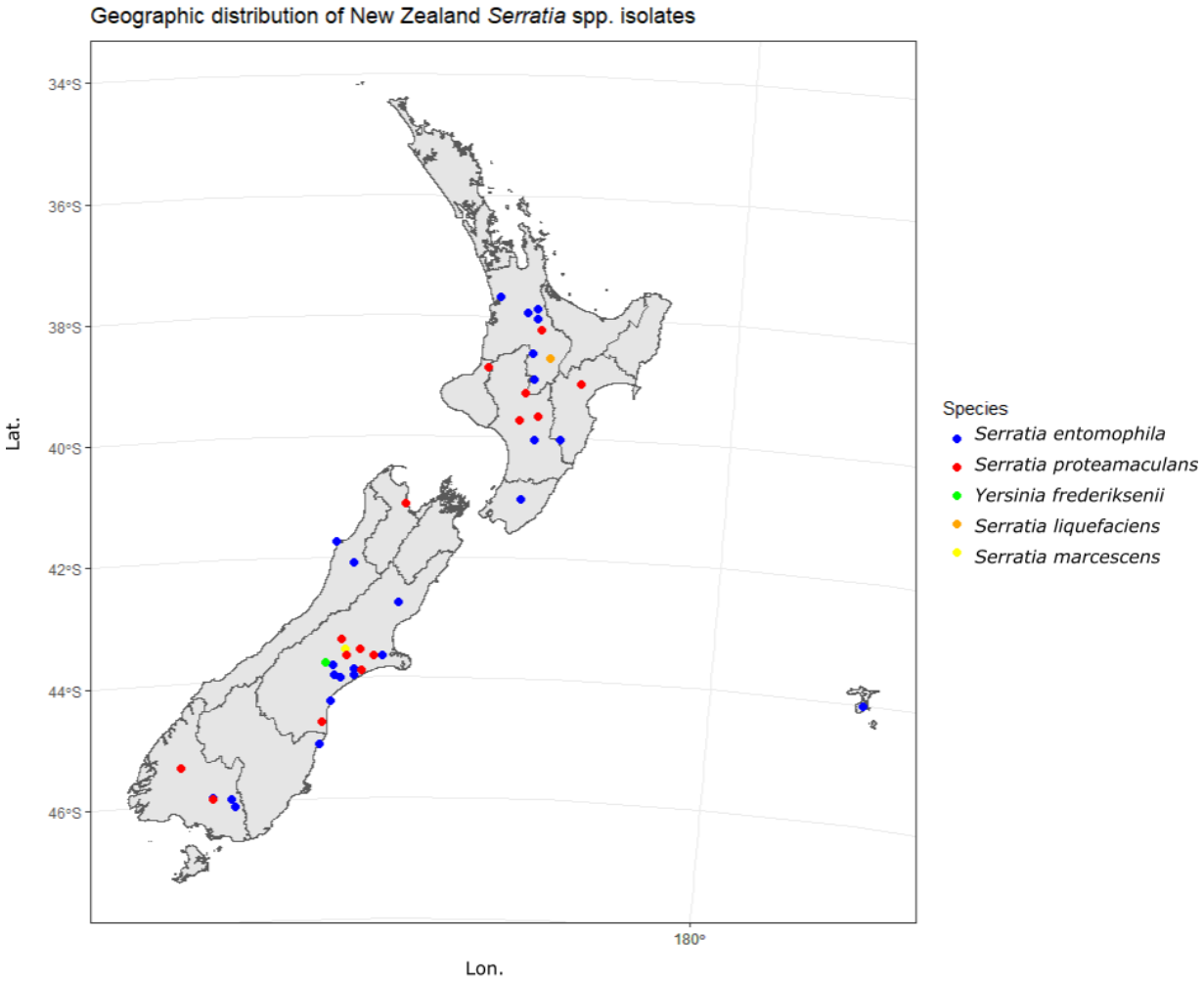


Figure 11 Geographic distribution *of Serratia* spp. used in this study across New Zealand
Frequency of isolates from a single location is not represented. Does not include *Serratia ficaria* isolate 457 or *Serratia entomophila* isolate 220 from France. Colour indicates species which is further explained within the key. Where possible, specific coordinates were utilised for isolation. For older samples where only site/ district was recorded, a central coordinate was assumed. Further details of each isolate can be found in Appendix B.1.

4.2.2 Bioassays of wildtype isolates against grass grub larvae

Before any phenotyping assays, gaps in the pathotype data of in-house isolates needed to be filled, and potential corrections to classifications that were determined from earlier research by Dodd (2003) were resolved. Disease type encompassed all disease phenotypes and not specifically just amber disease. Based on this a more robust classification system is shown in Table 9, where pathotype was defined as:

I. **Chronic** was defined as disease inflicted on 75% of treated larvae: Larvae with chronic infections will generally go on to have a longer disease time of up to 3-6 months which cannot be represented in a 12-day bioassay.

II. **Hyper** was defined as over 75% mortality with larval death occurring within 12 days of infection (standardised against the blank background control with a relevant P-value of <0.05).

III. **Non-pathogenic** was defined as less than 30% disease or no statistical difference from the background control samples.

Grass grub that had died from another means (e.g. accidental injury/ fungi) were removed from the total so as not to obscure the result. Of the pathotypes already defined in previous studies, only two (204 and 329) were altered with virulence observed in this study. Both 204 and 329 described as avirulent by Dodd (2003) are now defined as chronic. For the basis of this study, the most current pathotypes based on the aforementioned criteria will be used.

Table 9 Day 12-day bioassay assessment of selected *Serratia entomophila* and *Serratia proteamaculans* isolates.

P values (Fisher's exact), statistical significance in relation to the negative control highlighted in bold.

Strain	Disease + Std error (P-value) (%)		Mortality + Std error (P-value)		Combined + Std error (P-value)		Pathotype
Blank	0	±0	16.7	±11.23	16.7	±13.05	-
204[1]	58.3	±14.86 **(0.005)**	33.3	±8.33 (0.640)	91.7	±0 **(0.001)**	Chronic
207[1]	75	±13.05 **(<0.001)**	25	±0 (1.000)	100	±0 **(<0.001)**	Chronic
219[1]	0	±0 (1.000)	45.5	±15.74 (0.371)	45.5	±15.74 (0.193)	Non-path
220[1]	16.7	±11.23 (0.478)	25	±14.86 (1.000)	41.7	±14.86 (0.371)	Non-path
295[1]	0	±0 (1.000)	18.2	±12.19 (1.000)	18.2	±14.08 (1.000)	Non-path
305[1]	80	±13.33 **(<0.001)**	10	±10 (1.000)	90	±13.33 **(0.002)**	Chronic
306[1]	9.1	±9.09 (0.478)	45.5	±15.74 (0.371)	54.5	±14.08 (0.089)	Non-path
307[1]	30	±15.27 (0.078)	70	±0 **(0.027)**	100	±0 **(<0.001)**	Chronic
314[1]	8.3	±8.33 (1.000)	41.7	±15.07 (0.371)	50	±14.86 (0.193)	Non-path
315[1]	25	±13.05 (0.217)	8.3	±14.21 (1.000)	33.3	±13.05 (0.640)	Non-path
328[1]	0	±0 (1.000)	25	±13.05 (1.000)	25	±14.21 (1.000)	Non-path
329[1]	18.2	±12.19 (0.217)	63.6	±12.19 **(0.036)**	81.8	±12.19 **(0.003)**	Chronic
340[1]	70	±15.27 **(0.001)**	30	±0 (0.193)	100	±0 **(<0.001)**	Chronic
341[1]	66.7	±14.21**(0.001)**	25	±8.33 (1.000)	91.7	±0 **(0.001)**	Chronic
359[1]	66.7	±14.21 **(0.001)**	33.3	±0 (0.640)	100	±0 **(<0.001)**	Chronic
364[1]	50	±15.07 **(0.014)**	50	±0 (0.193)	100	±0 **(<0.001)**	Chronic
400[1]	0	±0 (1.000)	18.2	±12.19 (1.000)	18.2	±12.19 1.000)	Non-path
401[1]	0	±0 (1.000)	50	±15.07 (0.193)	50	±15.07 (0.193)	Non-path
438[1]	8.3	±8.33 (1.000)	33.3	±14.86 (0.640)	41.7	±14.86 (0.371)	Non-path
440[1]	0	±0 (1.000)	25	±13.05 (1.000)	25	±14.86 (1.000)	Non-path
442[1]	9.1	±9.09 (0.478)	18.2	±14.08 (1.000)	27.3	±15.74 (0.640)	Non-path
451[1]	66.7	±14.21 **(0.001)**	33.3	±0 (0.640)	100	±0 **(<0.001)**	Chronic
468[1]	72.7	±14.08 **(<0.001)**	9.1	±12.19 (1.000)	81.8	±0 (0.003)	Chronic
477[1]	8.3	±8.33 (1.000)	25	±14.21 (1.000)	33.3	±14.21 (0.640)	Non-path
480[1]	75	±13.05 **(<0.001)**	25	±0 (1.000)	100	±0 **(<0.001)**	Chronic
482[1]	8.3	±8.33 (1.000)	41.7	±15.07 (0.371)	50	±14.86 (0.193)	Non-path
485[1]	45.5	±15.74 **(0.014)**	54.5	±0 (0.089)	100	±0 **(<0.001)**	Chronic
562[1]	90.9	±6.3 **(<0.001)**	4.5	±4.5 (1.000)	95.5	±4.5 **(<0.001)**	Chronic
566[1]	8.3	±8.33 (1.000)	8.3	±11.23 (1.000)	16.7	±13.05(1.000)	Non-path
625[1]	0	±0 (1.000)	45.5	±15.74 (0.371)	45.5	±14.08 (0.193)	Non-path
766[1]	33.3	±14.21 (0.093)	0	±14.21 (1.000)	33.3	±15.07 (0.640)	Non-path
1021	0	±0 (1.000)	33.3	±14.21 (0.640)	33.3	±14.21 (0.640)	Non-path
1086[1]	9.1	±9.09 (0.478)	18.2	±14.08 (1.000)	27.3	±14.08 (0.640)	Non-path
1235[1]	9.1	±9.09 (0.478)	45.5	±15.74 (0.371)	54.5	±14.08 (0.089)	Non-path
1360[1]	9.1	±9.09 (0.478)	18.2	±14.08 (1.000)	27.3	±15.21 (0.640)	Non-path
1373[1]	8.3	±8.33 (1.000)	16.7	±13.05 (1.000)	25	±13.05 (1.000)	Non-path
1554[1]	8.3	±8.33 (1.000)	33.3	±14.86 (0.640)	41.7	±13.05 (0.371)	Non-path
1651[1]	90.9	±9.09 **(<0.001**)	0	±9.09 (1.000)	90.9	±9.09 **(0.001)**	Chronic
3041	0	±0 (1.000)	8.3	±8.33 (1.000)	8.3	±11.23 (1.000)	Non-path

[1] Indicates *S. entomophila* isolate

Based on the bioassay results of wildtype isolates, a table of isolates finalised to pathotype for each species could be constructed (Tale 10). Full pathotype data was then used for general comparisons with phenotypic assays o the secretion of accessory enzymes, such as chitinases and proteases, to determine any statistical difference between isolates of varying pathotypes. *S. proteamaculans* is the larger pathogenic sample set (be it chronic or hypervirulent) relative to non-pathogenic (Table 10). Of 34 *S. proteamaculans* isolates, 7 were non-pathogenic plasmid-free and 27 were pathogenic (either chronic or hypervirulent). The 56 isolates of *S. entomophila* included 32 non-pathogenic and 24 pathogenic isolates.

Results from these bioassays were then used in the studies reported in Chapters 5 and 6 to select further samples for *in silico* genome analysis so that both dominant pathotype traits (hyper and chronic) could be assessed in equivalent detail.

Table 10 Isolates of *Serratia* spp. and their assigned pathotype as determined in this study.
Pathotypes previously established or confirmed previously in Table 9
The number of isolates shown per species and pathotype is shown in brackets.
* Indicates isolate where pathotype is transient

Genus	Species		Pathotype		Isolate
Serratia	*entomophila*	(56)	Chronic	(24)	158, 176, 204, 207, 210, 305, 307, 340,341, 359, 364, 398, 451, 468, 480, 485, 562, 626, 1100, 1651, iDIA, A1M02, Sent1, Man3*
			Non-path	(32)	192, 219, 220, 294, 295, 306, 311, 314, 315, 328, 329, 345, 399, 400, 401, 438, 440, 442, 477, 482,498, 566, 625, 766, 1021, 1086, 1235, 1343, 1360, 1373, 1554, I
	proteamaculans	(34)	Chronic	(21)	L, M, RM5, 4, 142, 143, 145, 1048, CfB, 336, 1071, 1457,1769,12a, 12newD,20e, 25e, 28F, F, G, K
			Hyper	(6)	AGR96X, LC, 20093, 10novel, 1129, MH5
			Non-path	(7)	149, 299, 465, 495, 1137, 3041, Puna18
	marcescens	(1)	Non-path	(1)	1577
	liquefaciens	(3)	Non-path	(2)	377,376, F28
	ficaria	(1)	Non-path	(1)	457
Yersinia	*frederiksenii*	(1)	Non-path	(1)	Y49

4.2.3 BOX-fingerprinting

As described in Chapter 2, section 2.4.4, initial BOX-PCR bacterial fingerprinting with the BOX-A1R primer of the lettered isolates was undertaken by Amy Beattie (AgResearch, New Zealand) (unpublished data). These profiles demonstrated that *S. proteamaculans* has a diverse fingerprint profile, indicating a varied and diverse genome. Initial analysis of *S. entomophila* did not reveal any correlation between BOX PCR profile and pathogenicity. Genotyping comparisons were undertaken of A1M02 and its plasmid-cured isogenic isolate 5.6, which suggested the band differences shown are not due to the presence or absence of the pADAP plasmid (Figure 12). To begin to define any heterogeneity of *S. entomophila*, fingerprinting was undertaken across all isolates to determine the scale of variation. This method was also used to expand the *S. proteamaculans* dataset using those isolates that were not previously characterized in prior experimentation. Although there is a general trend of shared BOX-PCR profiles for some of the isolates of *S. proteamaculans* (Figure 13, panel D), the isolate profiles are non-uniform when compared to the *S. entomophila* A1M02 and *S. proteamaculans* AGR96X as references (Figure 13, panel D). Some isolates lack over half of the banding patterns observed in adjacent profiles, with the maximum of three homogenous isolate profiles (149, 336, 299) across the cohort of 17 isolates of *S. proteamaculans* (Figure 13, panel D). BOX-A1R amplification identified three main bacterial fingerprints shared between *S. entomophila* isolates (Figure 13 panels A-3). These three main profiles are characterized by two prevalent across both the North and South Islands, and a third BOX-PCR profile only found on the Chatham Island isolates and West Coast isolate MC2 (Figure 13 and Figure 15).

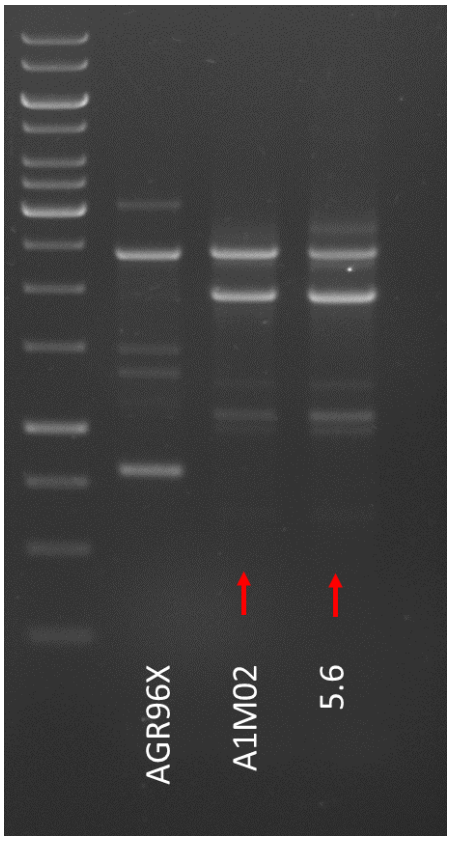


Figure 12 Agarose gel of genomic BOX PCR of *Serratia entomophila* plasmid free isogenic isolate 5.6 compared with WT plasmid carrying *A1M02 Serratia entomophila* and AGR96X used as a control.
Similar banding profiles of 5.6 and A1MO2 demonstrates that the presence of pADAP plasmid does not alter the genotype banding profile.

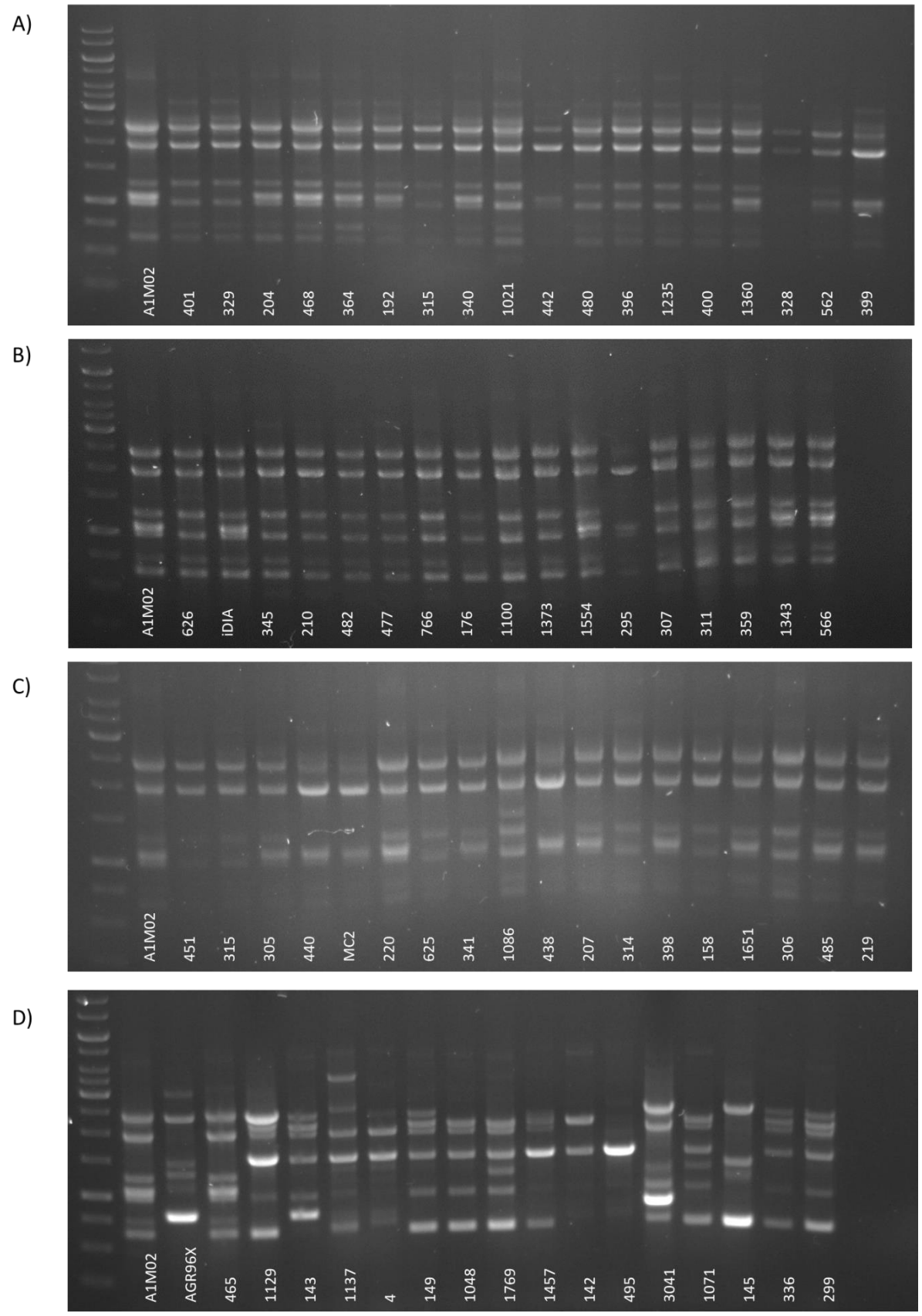


Figure 13 Agarose gel of genomic BOX fingerprinting of *Serratia entomophila*
Asterisk denotes *S. entomophila* isolate 626, currently used as biopesticide Invade®
Panels A-C show *S. entomophila* isolates whereas panel D) shows *S. proteamaculans.*

Considering the two prevalent profiles in New Zealand, analysis showed that Pasteur Institute isolate 220 (from France) has the same BOX profile (Figure 13, Panel C) as Canterbury isolates of *S. entomophila,* notably iDIA (chronic, Figure 13, Panel B). This indicates that these isolates are similar in origin (Figure 13). Referring to panel B of Figure 13, the *S. entomophila* pathogenic isolates 626 and iDIA, differ in their BOX PCR profile by one band in the central region indicating heterogeneity amongst chronic isolates from the same geographic region (Canterbury).

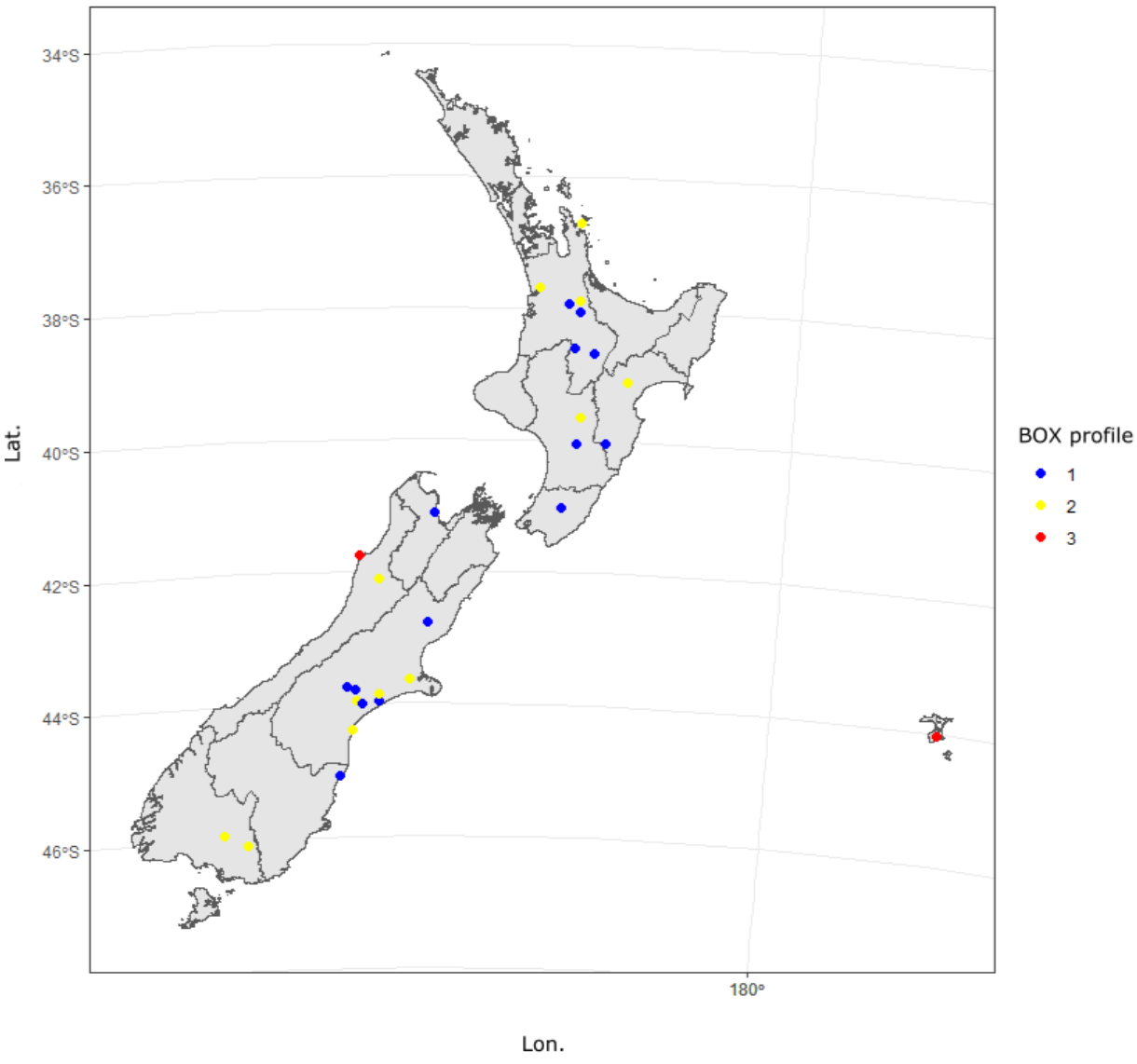


Figure 14 Geographical isolation of each isolate and their corresponding BOX fingerprinting profiles 1-3
Three BOX profiles shown with colour coordination in the key. BOX profile 3 was restricted to areas devoid of grass grub. Isolates only of New Zealand origin are shown. Isolate 220 from France not included, though shares the same BOX PCR profile as Canterbury isolate A1M02 (profile 2). Agarose gel of BOX profiles 1-3 shown in Figure 13.

BOX-profile three was found to be geographically specific to the Chatham Island and West Coast isolate MC2 (Figure 14). The only defining common factor between these regions is the lack of grass grub. These similarities warrant further investigation as both regions are geographically distant though share the same pathotype (profile cluster 3, non-pathogenic).

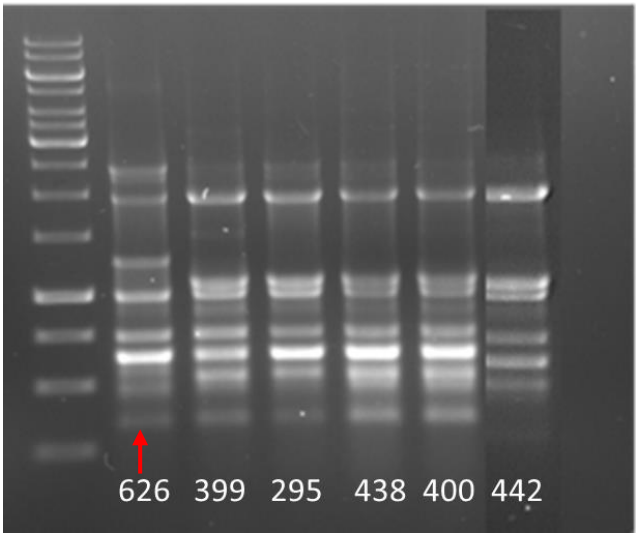


Figure 15 Agarose gel of genomic BOX bacterial fingerprinting of Chatham Islands isolates compared to *Serratia entomophila* 626 (red arrow).

In terms of *S. entomophila* vs. *S. proteamaculans*, the pathotype and geographic site of isolation of the isolates were compared to see if there is any consistency in the banding profile and pathovar/ geo-ecology. Only the *S. entomophila* Chatham Island isolates seem to have any geographic difference in fingerprinting profile (comparing Figure 13 panel A-D and Figure 15). The BOX profile of isolates from the North and South Islands are varied and overlapping with multiple isolates from Canterbury alone displaying one of the two variant profile types, although this could reflect movement due to differing microbial lifestyles.

BOX-fingerprinting of *S. proteamaculans* by Amy Beattie (unpublished data) revealed high degrees of diversity between isolates. Isolates do show similarity between one another; M and N have the same profile, showing a high level of similarity to K and 28F- though this is less than observed with *S. entomophila*. The hypervirulent isolates (AGR96X, LC, 20093, 10novel, 1129, MH5), though sharing similar pathogenicity and encoding AfpX, have diverse banding patterns amongst themselves.

4.2.4 Plasmid visualisation in *S. entomophila*

Through plasmid visualization using the method of Kado and Liu (1981), the presence and absence of pADAP was confirmed in isolates of both *S. entomophila* and *S. proteamaculans*. Isolates of *S. proteamaculans* had been previously assessed for the presence of pADAP plasmids by Dodd (2003) and Hurst (unpublished data). The results for *S. entomophila* assessed here for the presence of large plasmids are presented in Table 11 and Figure 16. The *S. entomophila* isolates A1M02 and the *S. proteamaculan*s isolate AGR96X were used as standard controls where A1M02 harbored a 153 Kb plasmid and AGR96X a 120 Kb plasmid. Non-pathogenic, plasmid-free isolates from the Chatham Islands have been screened and validated as pADAP-free by Dodd (2003) and re-confirmed in this study. Assessment of the Chatham Island isolates revealed a large plasmid (~120 Kb) in 438, 442, and 440. Chatham Island isolates 295 and 399 were found to be plasmid-free. No plasmid is found in West Coast isolate MC2, an isolate that shared the same BOX fingerprint profile to Chatham Island isolates (Figure 15).

Table 11 Mega plasmid visualisation of 56 isolates of *Serratia entomophila* both plasmid (+) and non-plasmid (-), and their relative size.
S. proteamaculans AGR96X used as a reference displayed in bold.

	plasmid visualised	Size (Kb)		Isolate
Serratia entomophila	+	153	(19)	158, 207, 210, 305, 307, 340, 345, 315, 359, 485, 562, 626, 1100, 1343, 1651, A1M02, iDIA, Man3, Sent1
		140	(1)	345[1]
		120	(5)	158, 345[1], 438,440, 442, 1021, **AGR96X**
		<120	(6)	158[1], 315, 438[1], 340[1], 566,
	_	Plasmid free	(29)	176, 192, 204, 219, 220, 295, 306, 311, 314, 328, 329, 341, 364, 399, 400, 401, 451, 477, 480, 482, 468, 625, 766, 1086, 1235, 1360, 1373, 1554, I

[1] Indicates duplicate names are shown when an isolate carries more than one plasmid of differing sizes.

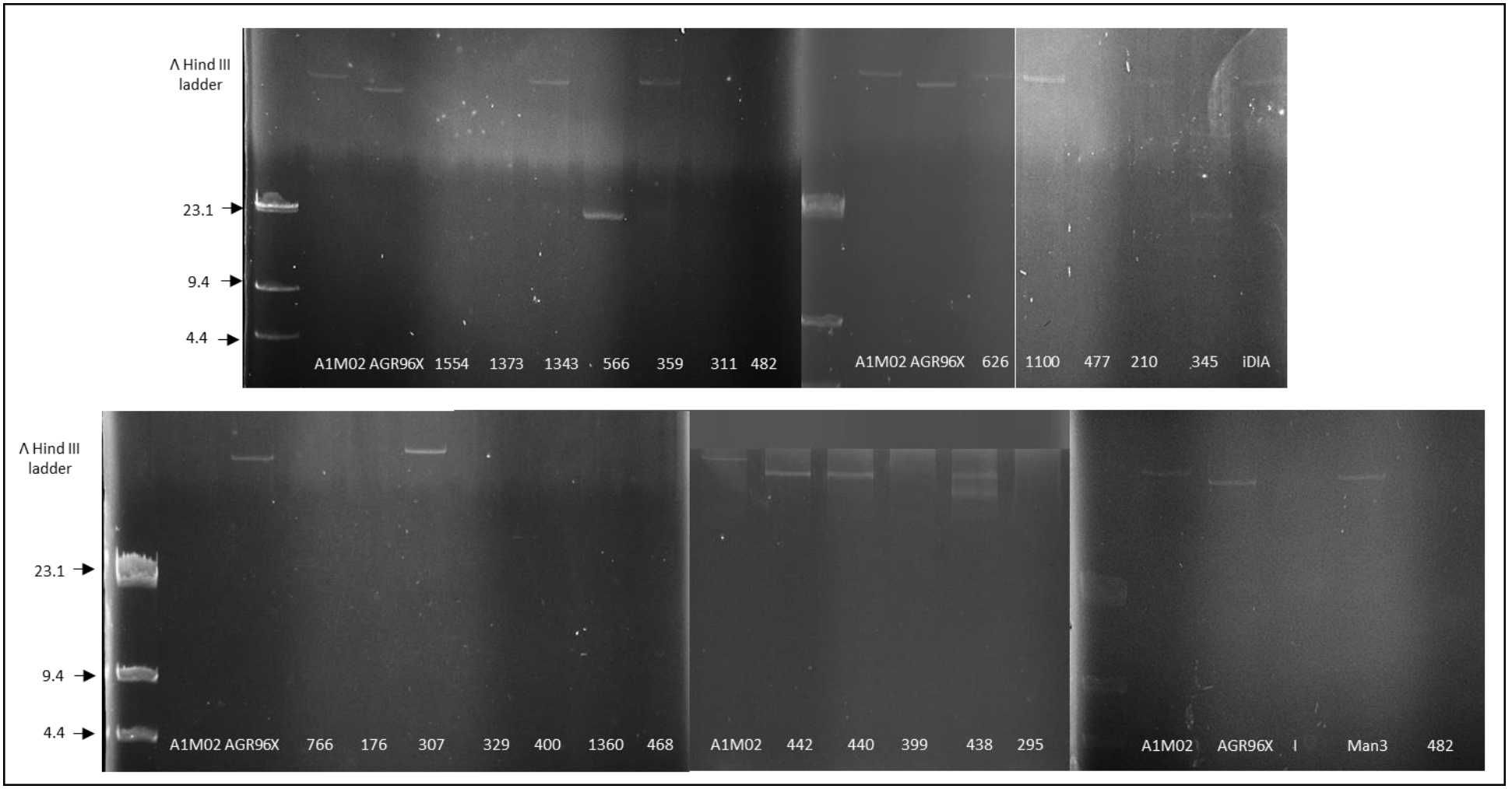


Figure 16 Kado and Liu mega plasmid visualisation of all isolates of *Serratia entomophila* using isolate A1M02 as a reference for the 153 Kb mega plasmid.

Serratia proteamaculans AGR96X used as a size reference for the smaller 120 kb pAFPX plasmid against Hind III λ ladder. Sizes shown are in Kb.

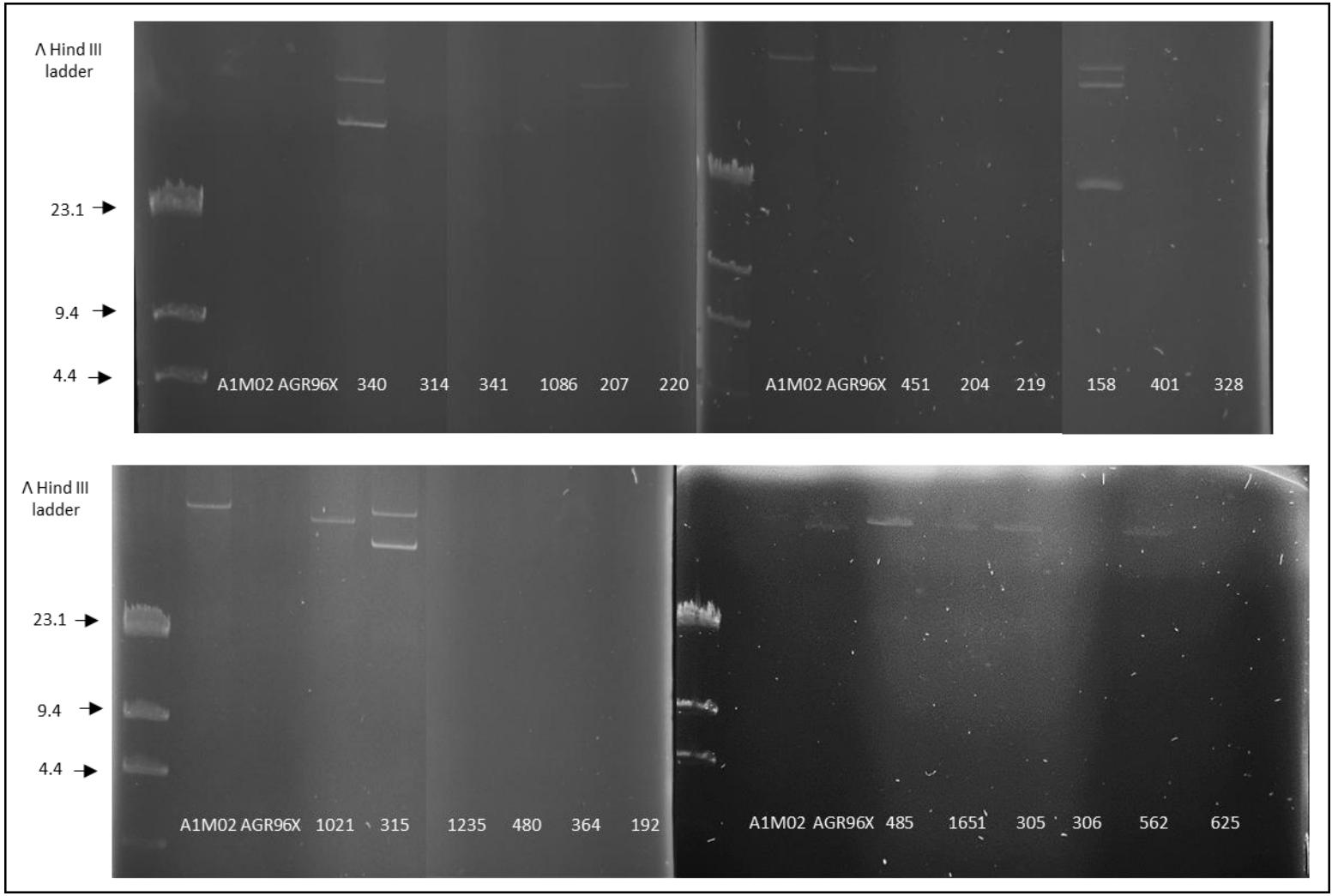


Figure 16 (continued) Kado and Liu mega plasmid visualisation of all isolates of *Serratia entomophila* using isolate A1M02 as a reference for the 153 Kb mega plasmid.
Serratia proteamaculans AGR96X used as a size reference for the smaller 120 kb pAFPX plasmid against Hind III λ ladder.

4.2.5 Screening for accessory virulence determinants

Using a range of plate assays (sections 4.2.5.1 to 4.2.6), each of the isolates was screened for accessory virulence determinants, i.e., proteases, lipases, chitinases, DNases, and siderophores that may reflect changes in virulence and/or ecological niche. The plate assays were prepared and analysed as outlined in Chapter 2, section 2.3.2.2 to section 2.3.2.8. Figure 17 shows the summarised mean results for all plate assays, split between species. The most variation between isolates was identified with levels of extracellular protease and lipase expression, alongside detection of the plant root stimulus auxin. Phosphatase expression was the most conserved across the two species, with a very narrow range of halo size, where an observable difference between the two species can be seen for DNase expression. This is expected due to the utilisation of DNase production as a species identifier between *S. entomophila* and *S. proteamaculans* and pertinent to the overarching hypotheses that it acquires less horizontally acquired DNA (O'Callaghan and Jackson 1993). For statistical comparisons, groupings were used as defined in Chapter 3 (Table 8). This included further splitting *S. proteamaculans* into the subspecies *S. proteamaculans* subsp. *proteamaculans* and *S. proteamaculans* subsp. *quinovora* as defined in section 3.2.

For each of the enzyme phenotyping assays below, interval plots were recorded using the ratios of the size of the halo clearing zones compared to colony size. This was undertaken for all replicates against species and pathotype with corresponding standard error bars. Data collected was then compiled and compared against known species and geographic collection site. A data matrix was then established compiling the geo-ecology and virulence expression phenotypic data (halo grouping) of all isolates, which can be found in Appendix B.3. Grouped interval plots were displayed as pathotypes broken down by species against halo ratio. Replicates for each isolate were used, and the mean of all grouped isolates and their replicates were used to construct standard error bars relevant for each group. AfpX encoding *S. proteamaculans* isolates (AGR96X, LC, 20093, MH5, and 1129) were presented in panel B of each enzyme phenotype result. Due to the inconsistent nature of assay data from hypervirulent isolates where large ranges of differentiation occurred between the isolates, results are presented as independent isolate values to display the differences observed in the phenotype of pAfpX encoding isolates.

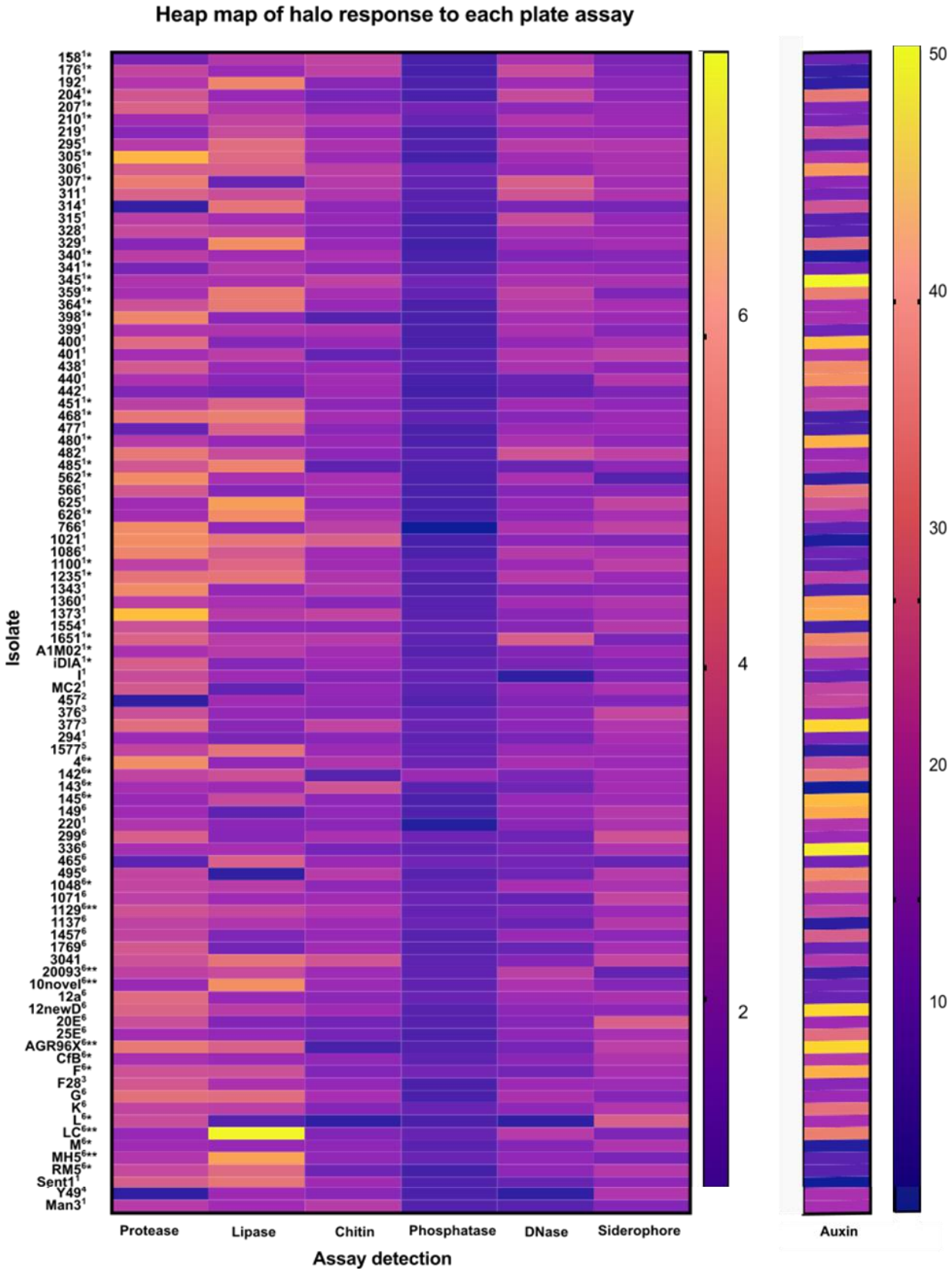


Figure 17 Summarised response of all isolates to plate assays for six treatments, and auxin detection 96 well assay of *Serratia*.

Plate assay response defined as the ratio of halo to colony size.

* Indicates chronic pathogen

** Indicates hypervirulent isolate

[1-6] Shows species. *Serratia entomophila* is marked with [1], and *Serratia proteamaculans* with [6].

4.2.5.1 Protease detection assay

The deviation of results across *S. entomophila* and *S. proteamaculans* for protease expression across pathotypes is negligible (Figure 18). Visual analysis of the interval plots revealed that between the two species the only non-overlapping standard errors were between pathogenic and non-pathogenic *S. entomophila* (Figure 18A). Otherwise, there was no obvious correlation between protease expression and pathotype. There was high heterogeneity in the response in hypervirulent classified isolates (Figure 18B).

To determine statistical significance, results were fitted to a general linear model before undergoing Tukey's range test between region, species, and pathotype. The resultant analysis of pathotype and species significance revealed that there was no statistical significance between the mean response of *S. entomophila* and *S. proteamaculans* (to 95% confidence P= 0.928) for the production of protease. When then separating further into subspecies as defined in Chapter 3, again there was no statistical difference in the groups assessed (*S entomophila* with *S. ficaria*, *S. proteamaculans* subsp. *quinovora* with *S. liquefaciens,* and *S. proteamaculans* subsp. *proteamaculans*). It can be observed as a trend that the differences between the *S. entomophila* group and *S. proteamaculans* subsp*. quinovora* (P=0.425) differ more than against *S. proteamaculans* subsp. *proteamaculans* (P=0.804).

When looking at pathotype, again, no statistical significance was observed with the relationship between pathogenic and hypervirulent isolates having a P-value of 0.073, although a trend can be observed in *S. entomophila* where non-pathogenic isolates show an overall reduced production of proteases.

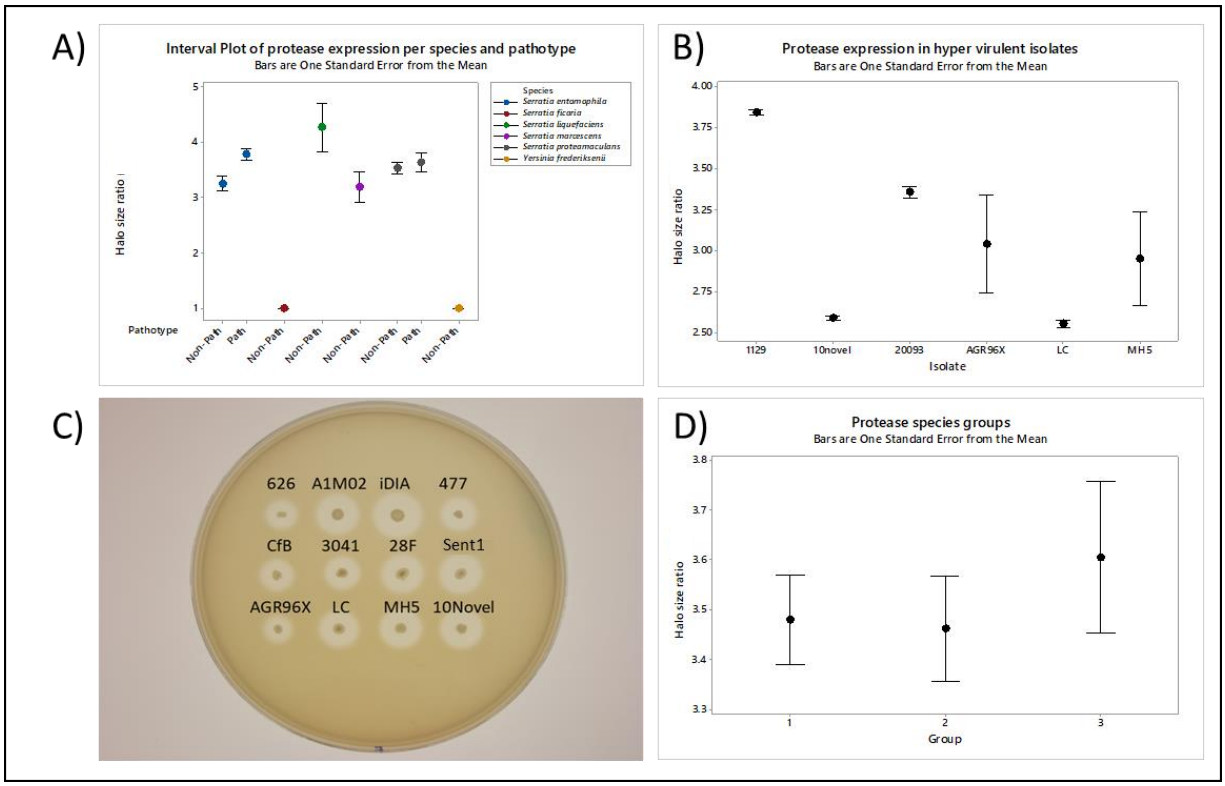


Figure 18 Results of protease plate assays across all *Serratia entomophila* and *Serratia proteamaculans* isolates
A) Average ratio of protease expression with standard error for all isolates split by species and pathotype B) Individual results of the mean for hypervirulent isolates with standard error C) example plate showing (top row) *S. entomophila* (middle row) *S. proteamaculans* and (bottom row) hypervirulent isolates D) shows average plate response when separated into subspecies groups defined in Chapter 3.

4.2.5.2 Lipase detection assay

Assessment of lipase production revealed a similar response as protease (Figure 19). Both *S. entomophila* and *S. proteamaculans* show markedly higher production of lipases for pathogenic isolates in comparison to non-pathogenic, specifically with isolate LC (Figure 19B), although this seems to be more defined as a trend than with protease expression. Generally, the lowest production was for non-pathogenic *S. proteamaculans*. For hypervirulent isolates, there were large differences in protease production between isolates. However, lipase production by MH5 and LC was higher than other hypervirulent isolates, including AGR96X, which was quite poor in production (Figure 19B). Comparative analysis of all isolates shows no statistical significance between either species (P= 0.30) or either between chronic or non-pathogenic isolates (P= 0.092) (Figure 19A). When using delineations of subspecies and groups (Table 8), again no statistical difference was found between any group (Figure 19D). However, significant differences were observed between hypervirulent isolates and both chronic

(P= 0.001) and non-pathogenic (P <0.001) isolates. As with protease production, a trend can be observed in *S. entomophila* where non-pathogenic isolates show an overall reduced production of lipases.

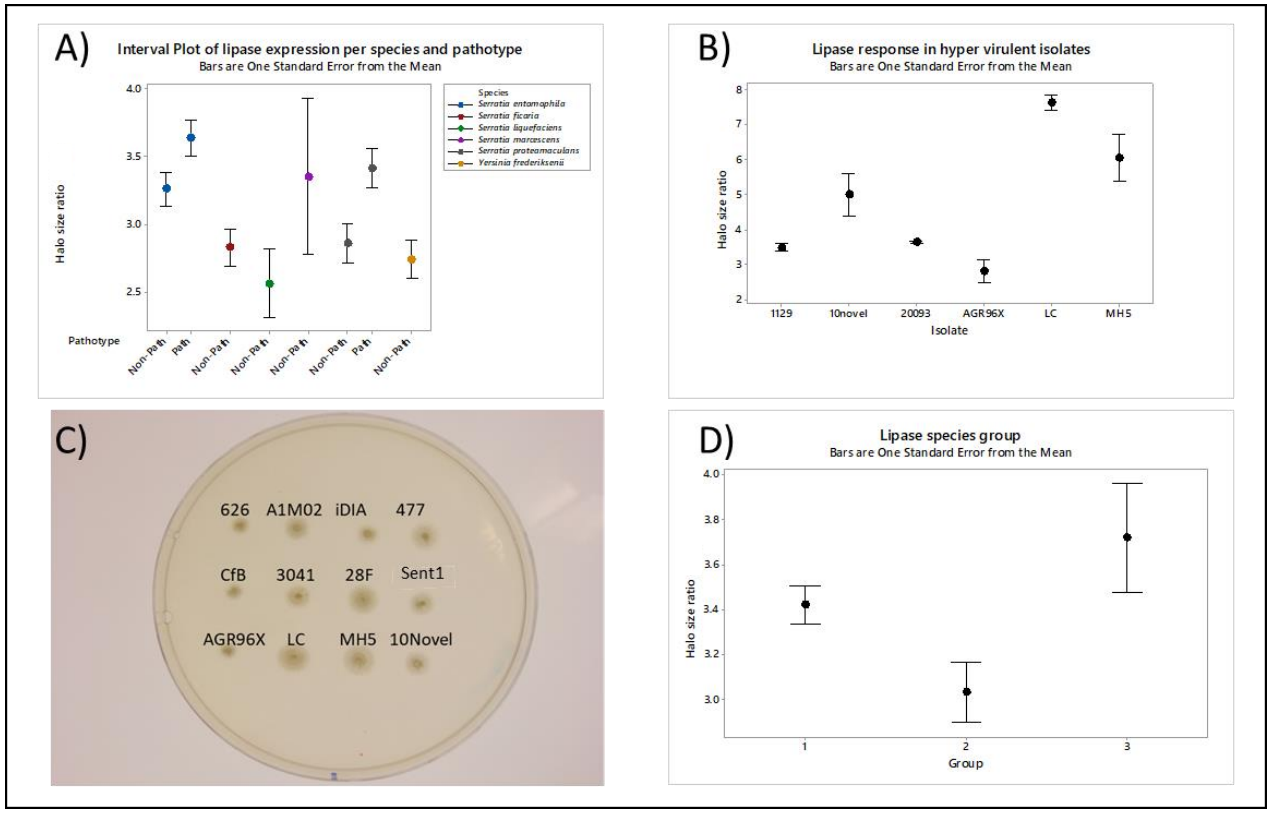


Figure 19 Results of lipase plate assays across all *Serratia entomophila* and *Serratia proteamaculans* isolates
A) Average ratio of lipase expression with standard error for all isolates split by species and pathotype B) Individual results of the mean for hypervirulent isolates with standard error C) example plate showing (top row) *S. entomophila* (middle row) *S. proteamaculans* and (bottom row) hypervirulent isolates D) shows average plate response when separated into subspecies groups defined in Chapter 3.

4.2.5.3 Chitinase degradation assay

Screening isolates for chitinase production (Figure 20) revealed an inverse effect relative to either protease or lipase. Non-pathogenic strains, on average, had higher chitinase activity. However, the standard error bars overlap for pathogenic isolates for both species. Among all *Serratia* spp., *S. liquefaciens* and *S. marcescens* had the most variation of chitinase production, whereas *S. entomophila* and *S. proteamaculans* had smaller ranges (Figure 20A). Again, the comparative analysis showed no significant difference between either species or pathotype. AGR96X and LC are unique in their group for expression (Figure 20B), with on average lower chitin degradation. General linear model analysis

revealed no statistically significant differences between the means of any species, with *S. entomophila* to *S. proteamaculans* comparisons significant to a P-value of 0.629. The same result was acquired when using subspecies groupings previously defined (Figure 20D), though the lowest difference was observed between the two subspecies of *S. proteamaculans* (P=1.000). The same result was observed for all three pathotypes (hypervirulent, chronic, and non-pathogenic), where differences between non-pathogenic and pathogenic (chronic or hyper show no significance (P=0.461 and P= 0.997 respectively). A trend between chronic and non-pathogenic isolate can be observed in *S. entomophila* in Figure 20A where higher chitinase production can be seen in the non-pathogenic isolates. This conflicts with the previous observations in lipases and proteases, where results showed the pathogen to produce higher levels of accessory virulence factors.

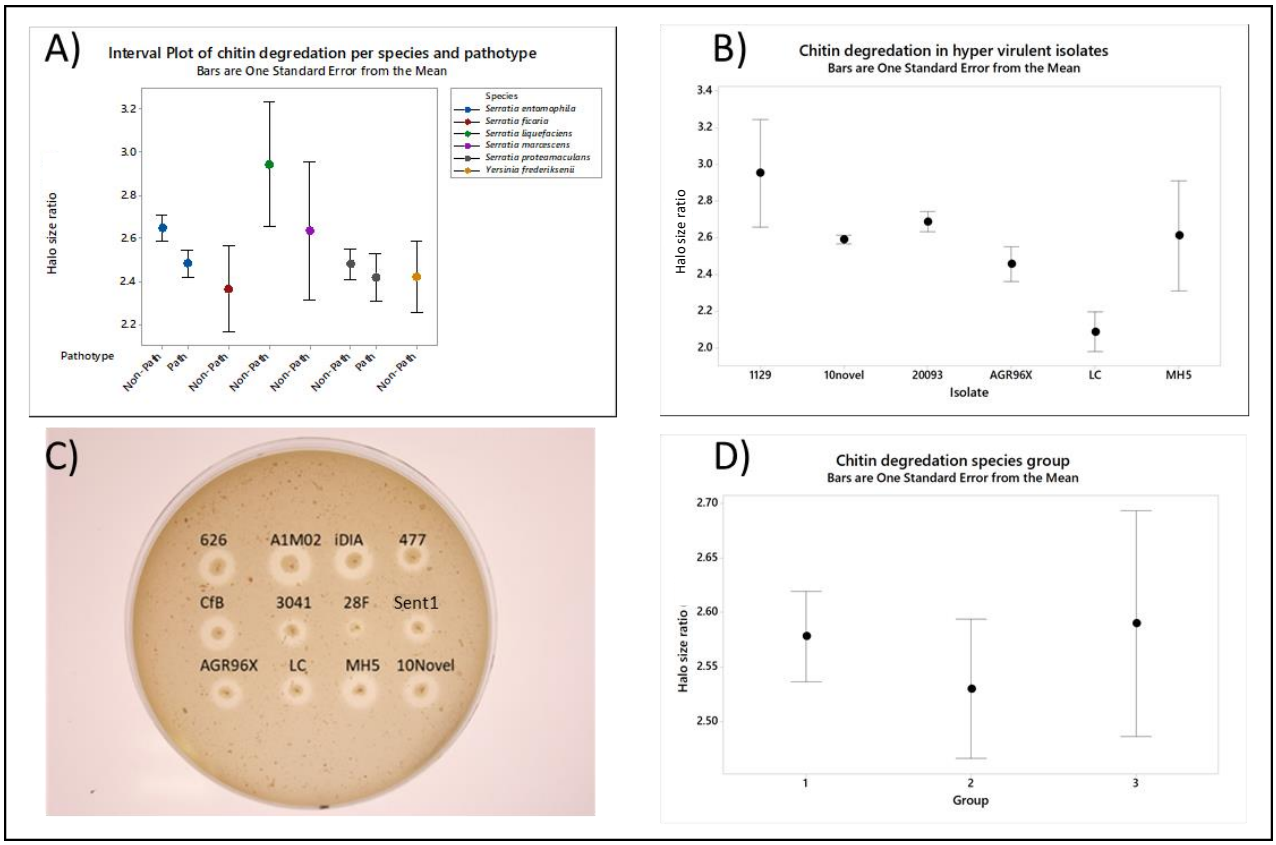


Figure 20 Results of chitin plate assays across all *Serratia entomophila* and *Serratia proteamaculans* isolates

A) Average ratio of chitinase expression with standard error for all isolates split by species and pathotype B) Individual results of the mean for hypervirulent isolates with standard error C) example plate showing (top row) *Serratia entomophila* (middle row) *Serratia proteamaculans* and (bottom row) hypervirulent isolates D) shows average plate response when separated into subspecies groups defined in Chapter 3.

4.2.5.4 Siderophore production

Tukey's comparisons for siderophore production revealed that hypervirulent isolate expression was, on average, significantly different from the chronic and non-pathogenic isolates assessed (Figure 21). On average, these were lower in expression than both standard pathotypes of *S. proteamaculans* and the non-pathogenic *S. entomophila* but showed similarities with the chronic isolates (Figure 21A). Although marginally different in mean expression, neither species was statistically different from one another. As with previous trends, there was no discernible significant difference between the means of response for any species. The smallest P-value regarding species comparison was between *S. entomophila* and *S. proteamaculans*, at P= 0.195. When observing at the subspecies level again no significant difference was identified. A general trend showed the means of response between *S. entomophila* and *S. proteamaculans* subsp. *quinovora* differed higher than any other comparison (P=0.068). *S. entomophila* has the lowest average siderophore production of all three groups (Figure 21D). At the pathotype level, hypervirulent was again significantly different in comparison to the other two pathotypes. P-value comparison related to chronic was P= 0.030, whilst to non-pathogenic isolates, it was P=0.002.

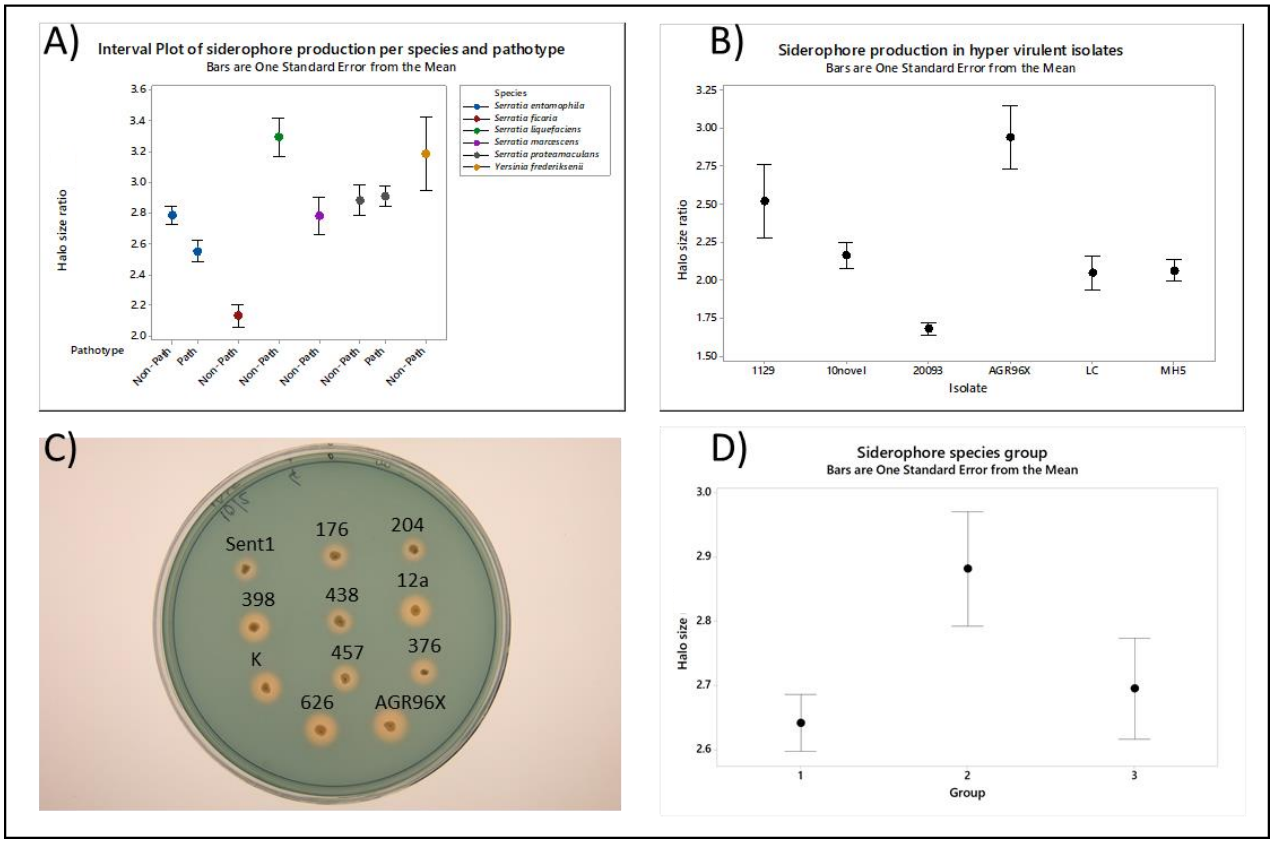


Figure 21 Results of siderophore plate assays across all *Serratia entomophila* and *Serratia proteamaculans* isolates

A) Average ratio of siderophore expression with standard error for all isolates split by species and pathotype B) Individual results of the mean for hypervirulent isolates with standard error C) example plate showing (top row) *Serratia entomophila* (middle row) *Serratia proteamaculans* and (bottom row) hypervirulent isolates D) shows average plate response when separated into subspecies groups defined in Chapter 3.

4.2.5.5 DNA hydrolysis

In contrast to the phenotypes observed in lipase and protease assays, assessment of DNase production showed a significant difference (P<0.001) in halo formation between both *S. entomophila* (mean = 2.9 mm) and *S. proteamaculans* (mean = 2.3 mm) (Figure 22A). There were also significant differences observed for non-pathogenic isolates compared to both chronic and hypervirulent pathotypes (P= 0.003 to chronic as observed in Figure 22A, and P= 0.001 with hypervirulent) (Figure 22A and B). Interval plots showed these results clearly, where *S. entomophila,* on average, had larger halo ratios, with smaller deviations. The small error bars indicate a conserved phenomenon across the population. This is again reiterated when the isolates were split into the subspecies grouping. *S. entomophila* was statistically the most responsive, whereas the means of both subspecies of *S. proteamaculans* did not differ much in average halo size (0.017 mm average difference, P=0.991) (Figure 22D).

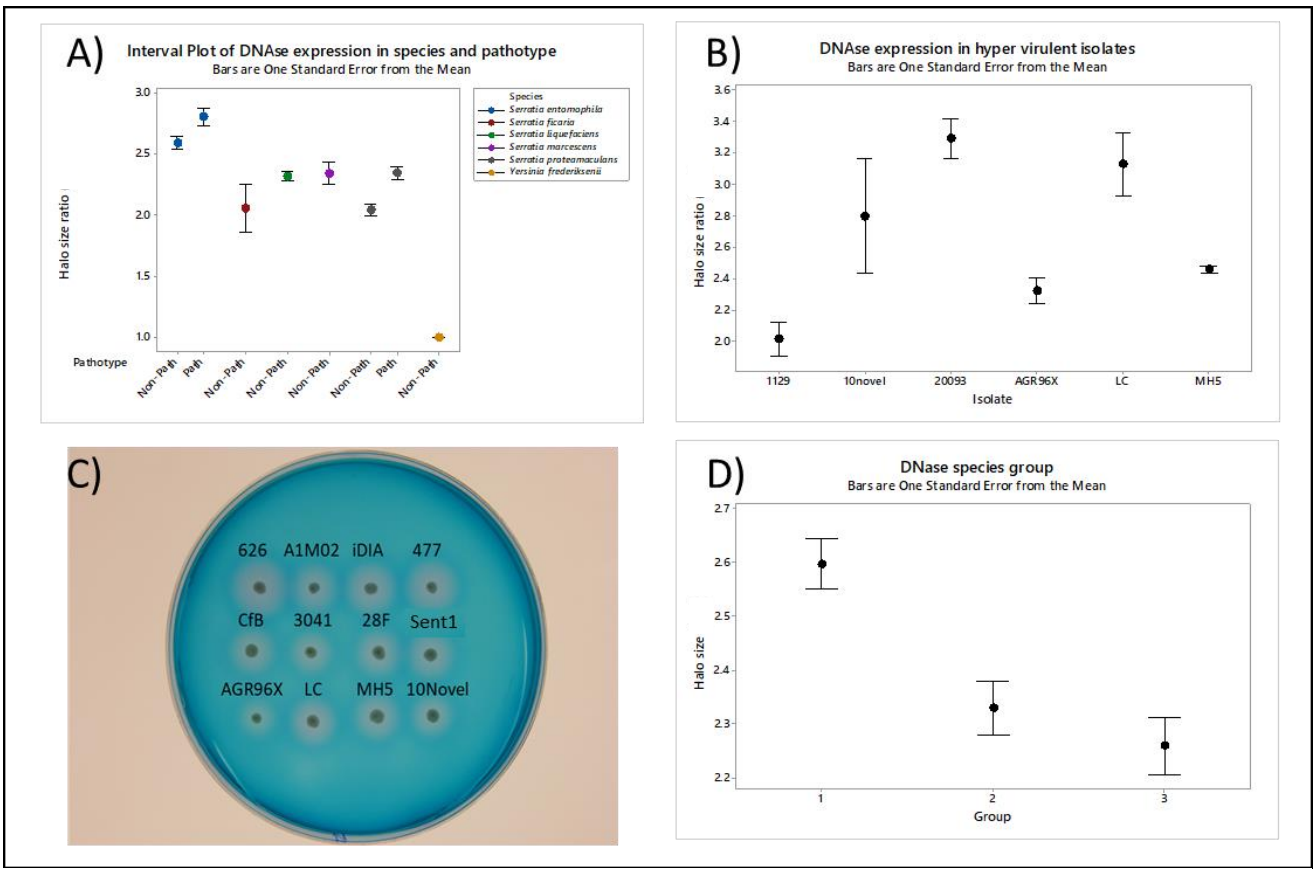


Figure 22 Results of DNase plate assays across all *Serratia entomophila* and *Serratia proteamaculans* isolates

A) Average ratio of DNase expression with standard error for all isolates split by species and pathotype B) Individual results of the mean for hypervirulent isolates with standard error C) example plate showing (top row) *S. entomophila* (middle row) *S. proteamaculans* and (bottom row) hypervirulent isolates. Note: *S. proteamaculans* isolates on average are smaller in halo size than *S. entomophila* D) shows average plate response when separated into subspecies groups defined in Chapter 3.

4.2.6 Screening for traits for plant growth-promotion

As the majority of *Serratia* spp. assessed in this study were isolated from soil or grass grub larval macerates, all 95 isolates were assessed for expression of plant growth-promoting traits to identify any such factors that could potentially place plants or plant roots as an ecological niche. Thus, auxin and phosphate solubilising media were selected as indicators.

Auxin expression was assessed with 96 well plate assays using Salkowski's reagent to identify indole-3-acetic acid (IAA) in the precipitate of cultured isolates (Chapter 2, section 2.3.3). No statistical significance was observed between the mean response of any species, including between the two species of interest based on the calibration curve shown in Figure 23A (P=0.996) (Figure 23B). The means for both *S. entomophila* and *S. proteamaculans* overlap quite closely for all pathotypes. Including the subspecies groupings (Table 8) there were still no observably statistical differences, though again as previously, a trend is observed where *S. entomophila* and *S. proteamaculans* subsp. *quinovora* have the furthest difference between means (4.81 mg/µL difference, P=0.167) (Figure 23E). Non-pathogenic to pathogenic isolates gave P values of 0.996 and 0.977, whilst hyper to chronic was P=0.969 (Figure 23B).

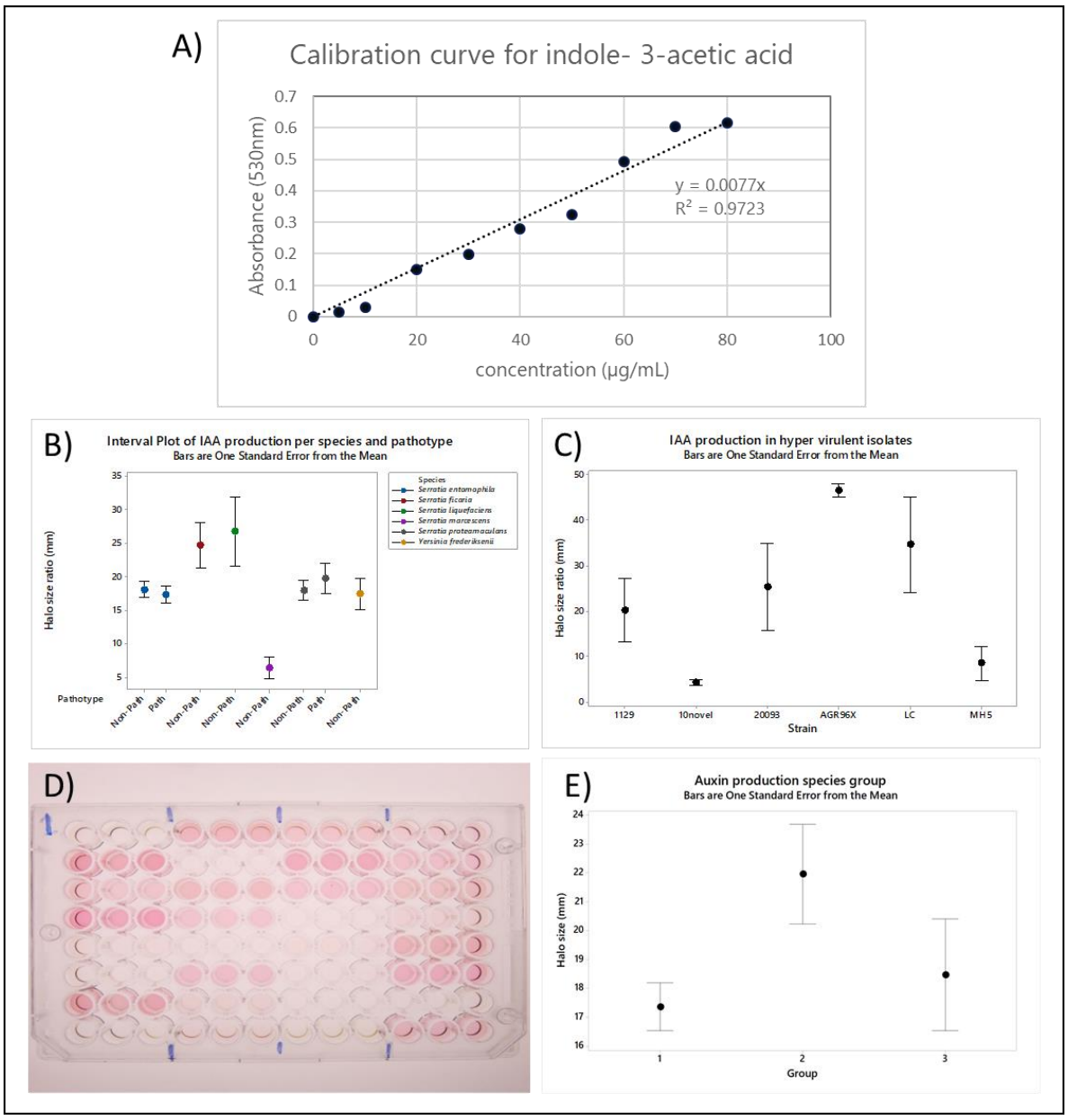


Figure 23 Results of auxin 96 well assays across all *Serratia entomophila* and *Serratia proteamaculans* isolates
A) Calibration curve for absorbance reading at 540nm of concentrations of IAA and Salkowski's reagent B) Average ratio of IAA expression with standard error for all isolates split by species and pathotype C) Individual results of the mean for hypervirulent isolates with standard error. D) example 96-well plate E)) shows average plate response when separated into subspecies groups defined in Chapter 3.

Secondly, to determine relative rates of phosphate solubilisation, which increases soluble phosphorus to plant root systems, Pikovskaya's plate assays were used (Chapter 2, section 2.3.2.8). Assessment of the resultant phenotypes revealed that, on average, *S. proteamaculans* has a larger halo to colony ratio, implying the increased presence of phosphatases secreted (P= 0.002) (Figure 24). This was the

only significant observation, as no difference was observed for pathotypes amongst species and no significant difference was observed between any of the pathotypes. This is again supported when using the subspecies grouping. No difference is observed between both subspecies of *S. proteamaculans* (P=0.458), whist *S. entomophila* was significantly different from both (P=0.003 and P<0.001 respectively). As previously observed, trends in *S entomophila* show that non-pathogenic isolates solubilise phosphates to a lower degree than pathogens.

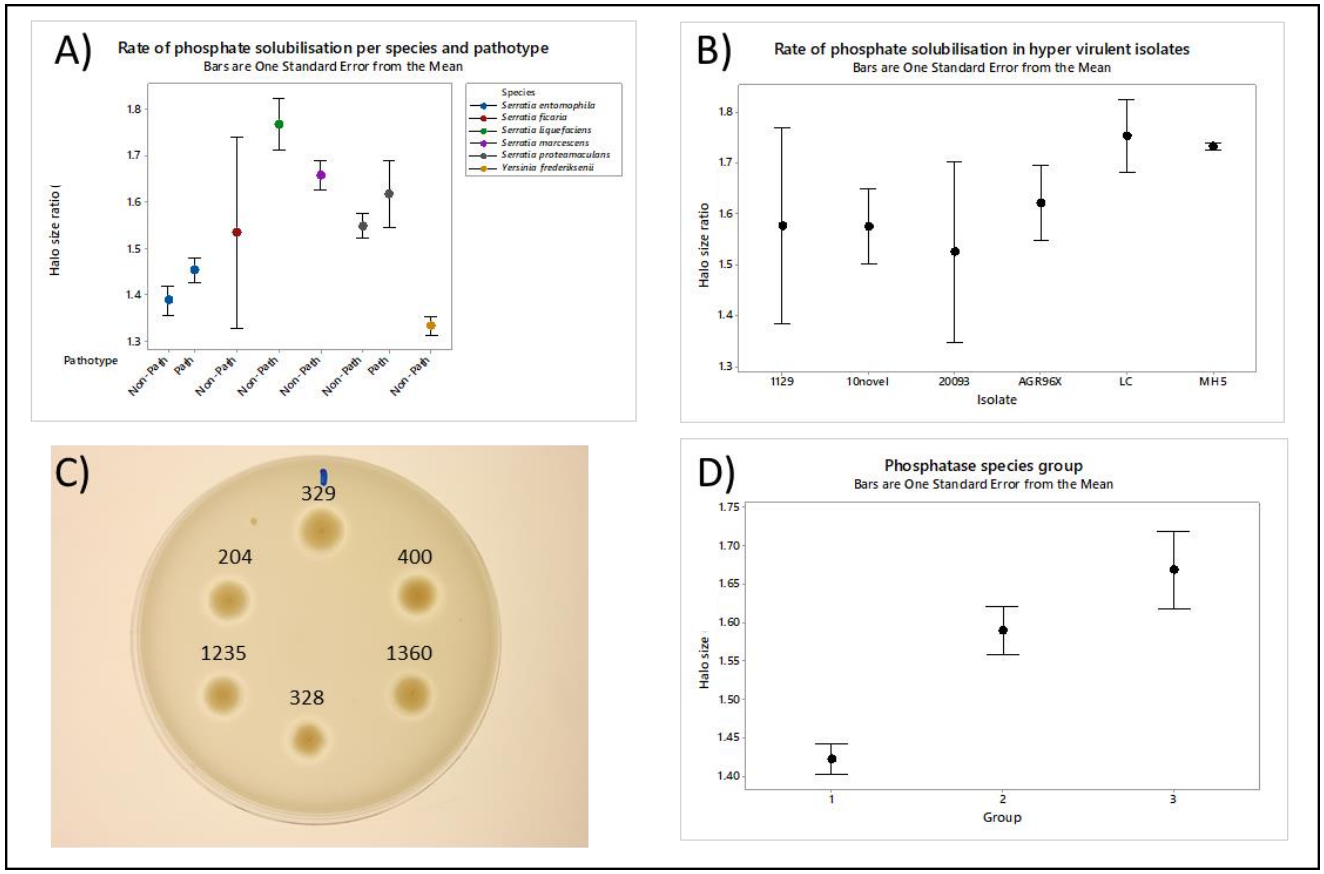


Figure 24 Results of phosphatase plate assays across all *Serratia entomophila* and *Serratia proteamaculans* isolates

A) Average ratio of phosphatase expression with standard error for all isolates split by species and pathotype B) Individual results of the mean for hypervirulent isolates with standard error C) example plate showing (top row) *S. entomophila* (middle row) *S. proteamaculans* and (bottom row) hypervirulent isolates.

4.3 Discussion

BOX bacterial fingerprinting established profile patterns in 56 *S. entomophila* isolates belonging to two pathotypes (chronic and non-pathogenic, section 4.2.3). Three BOX clusters were identified, although none of these corresponded with pathotype when compared with the results from previous and current bioassays. This concurs with and expands on the results found by Claus et al. (1995), where two biotypes of *S. entomophila* profile were identified by BOX bacterial fingerprinting, but no link to pathotype could be established.

In addition to pathotype, genotyping differences were assessed between geographic sources. Two of the clusters identified in this study were sourced from overlapping regional areas across the South Island of New Zealand, while one cluster was geographically unique, being solely based on isolates from areas that are devoid of grass grub (the Chatham Islands and the West Coast of the South Island). These regions are geographically distant, at over 1,000 km of separation from the South Island of New Zealand. It should be acknowledged that not all West Coast Isolates share a similar pattern. Isolate 219 from Reefton (centrally located between the West and East coast) aligns more closely with one of the larger clusters of fingerprinting profiles, though the geographic isolation point of this sample is not as distinct as MC2. Possible explanations for the distribution of these geographically widespread groups can be from a range of mechanisms such as the migratory paths of birds within New Zealand, which are known predators of grass grub. The white-fronted tern (tara) population of the Chatham Islands has been well recorded as dispersing from Autumn to Australasia, and frequently found along the New Zealand coastlines (Aikman and Miskelly 2004). The dissemination of soil microbes by the movement of people and agricultural machinery across the North and South Islands would play a significant role in the dispersal of geographically isolated bacteria. The application of the biocontrol agents Invade™ and Bio-Shield® has been widespread across New Zealand. Dispersal of biological bacterial controls allows the dissemination of evolutionarily similar organisms across New Zealand, used in their production. In this case, the initial production of Invade™ utilised isolate A1M02 before replacement with the phage-resistant isolate 626 under the production of Bioshield®(Johnson et al. 2001). Both isolates have differing BOX fingerprinting profiles (Figure 13) as either cluster one or cluster two. Further analysis of isolation dates of *S. entomophila* used in this study show that prior to widespread use of 626 as a biocontrol agent, similar distributions of profiles were seen. This would support that the genotyping was a fair representation of the wild *S. entomophila* population and not affected by application of Invade™ or Bio-Shield®. Tian et al. (2018) found a close link between bacterial communities in soils of differing soil organic matter decomposition components and climate, with 80.5% of variation structure correlated to geographic locale and environmental factors. As New Zealand has such a varied and dynamic topography alongside the diverse land and soil use types, it can be assumed that this will have a large impact on the structure of bacterial communities. Alongside the

prevailing wind of New Zealand coming from a north-west aspect, the wet weather that accumulates on the West Coast of the South Island of New Zealand is thought to prevent grass grub establishment, providing another link between climate and ecosystem.

The variability in *S. proteamaculans* genotype observed in this study and previous research shows that *S. proteamaculans* are more genetically diverse, which could reflect the diversity we see in *S. proteamaculans* pathotypes. *S. proteamaculans* is not currently used as a biocontrol agent. Which could reflect the diversity of geographically isolated bacteria, as wide dispersal of a uniform isolate profile has not been undertaken. Further analysis in Chapter 6 will characterise the chromosomes of *S. proteamaculans* to expand on the differences in fingerprinting profiles.

Through the use of various plate-based assays, the expression of accessory virulence factors differed between *S. entomophila* and *S. proteamaculans*. Amongst the enzymatic assays, protease, chitinase, and auxin detection assays showed no statistical difference between these species or their pathotype. DNase production was notably different for both pathotype and species. *S. entomophila* showed a higher presence of DNases than *S. proteamaculans* and non-pathogenic isolates also outscored pathogenic isolates. Detection of reduced production of DNases in *S. proteamaculans* isolates compare to that of *S. entomophila*, and pathogenic isolates of *S. entomophila* compared to non-pathogenic con-specifics are interesting in the context of the aims of this study. *Serratia marcescens* produces high levels of extracellular nucleases (Suh et al. 1996). It has been observed that higher levels of extracellular DNase secretion by bacteria can limit the transfer of DNA by transformation (Blokesch and Schoolnik 2008). Studies conducted by Focareta and Manning (1991) assessing *Vibrio cholerae*, produced DNase deficient mutants that were able to take up plasmid DNA by transformation readily. This was as opposed to the transformation limiting wildtype isolates, with no limiting effect on virulence compared to DNase secreting wildtype. Based on this it is plausible that the higher secretion levels of extracellular DNases in *S. entomophila* may reduce the uptake of free DNA from the environment, therefore limiting genetic evolution as a bottleneck relative to other localised bacteria. This, in turn, will limit variation and thus, a reduced capacity of *S. entomophila* to increase virulence. This scenario in part helps answer the overarching hypothesis of the study, that we can accept *S. entomophila* is more genetically conserved than *S. proteamaculans*. Wildtype bioassays show each pathotype of *S. entomophila* is less varied in phenotype than *S. proteamaculans*, causing a chronic amber disease state or no pathogenicity at all. Limited genomic evolution, facilitated by various genetically driven mechanisms including DNase secretion, could have limited acquisition of virulence associated genes thereby limiting disease evolution. Further analysis, including *in silico* genome comparisons in Chapters 5 and 6, will seek to clarify whether any other mechanisms can be identified and any, or what, limitations to the genome this has resulted in in comparison to *S. proteamaculans.*

Lipolytic enzyme secretion has been extensively studied in clinically important bacteria such as *Candida albicans*, where it has been implicated in bacterial colonisation and cytotoxicity (Park et al. 2013). Phospholipases, in particular, are known to catalyse the hydrolysis of phospholipids which are a key component of cell membranes of the gastrointestine of higher eukaryotes (Flores-Díaz et al. 2016). The lipase production observed in the assessed *Serratia* spp. aligns with the trend we see where pathogenic isolates produce more lipases than non-pathogenic. Interestingly, the fact that hypervirulent isolates of *S. proteamaculans* have a higher level of production to 99% confidence reflects their speed of the invasion and disease process. This is consistent with the study of Hurst et al. (2018) which reported that within 24-72 h post ingestion, larvae challenged with AGR96X became discoloured and increased mortality was recorded over 12-days (87.5 ± 4.5%).

As outlined in section 4.2.5.4, siderophore production, which is required for capturing trace elements affording the bacterium a competitive advantage in both the environment and the insect host, differed only between isolates of differing pathogenicity's (potentially a recognisable marker of pathogenicity). Siderophore production has been implicated in virulence in other gram-negative organisms, such as the staphylobactin production of *Staphylococcus aureus* (Dale et al. 2004). Cessation of production of staphylobactin in *S. aureus* has been demonstrated by Dale et al. (2004) to result in an attenuated pathogen. In *Serratia,* the heightened production of siderophores in pathogenic bacterial isolates in comparison to non-pathogenic isolates also implicates these compounds as virulence factors. This supports one of the main hypotheses of this study, (Chapter 1, hypotheses v), that pathotype will influence the detection of accessory virulence determinants.

Assessment of plant growth-promoting traits such as auxin production and phosphate solubilisation was assumed to be related to benefits to a bacterial species or pathotype in niche establishment. Auxin production in *Salmonella* colonising seedlings as a secondary infection showed an increase in secondary root formation (Cox et al. 2018). No difference was seen in this study between *S. entomophila* and *S. proteamaculans* in their auxin production, but phosphate degradation showed a trend of higher solubilisation in *S. proteamaculans* isolates. Phosphate solubilising bacteria are known to hydrolyse phosphorus compounds to a form that can be used by plants for generating ATP (Kalayu 2019). By doing this, a bacterium can facilitate the growth and optimisation of a niche by facilitating plant health and optimising the ecological niche for an insect host. In relation to *S. entomophila* and *S. proteamaculans,* less solubilisation by *S. entomophila* could support the hypothesis of a reducing genome moving towards an endopathogenic relationship with *C. giveni*. In other words, a reduced disease state maintains the grass grub host and provides a long term niche for *S. entomophila*. This would support the hypothesis of *S. entomophila* having a conserved pathotype (Chapter 1, hypothesis

i) and potentially explain why populations of *S. entomophila* collapse after reduction in *C. giveni* larvae populations. On the other hand, *S. proteamaculans* secreting more of these phosphatase enzymes potentially indicates that the presence of *S. proteamaculans* could benefit plant growth and therefore enhance the bacterium's capacity to be selected for the rhizosphere and colonisation of the root system. In the instance of *S. proteamaculans*, *C. giveni* may not be its preferred environmental niche. The degree of pathogenicity to the host is less important for bacterial survival, potentially mediating the varied pathotype we see throughout the species across both subspecies.

In summary, assessment of *S entomophila* and *S. proteamaculans* accessory virulence determinants was, until now, undefined. Here the presence or absence of accessory virulence determinants associated with pathotype was elucidated. Isolates of *Serratia proteamaculans* are visibly more varied in genotype, plasmid type, and accessory virulent determinants than isolates of *S. entomophila.* Pathotype shows distinct variation when comparing hypervirulence to classic pathotype, but geographic isolation site seems to show nearly no effect on trait acquisition, except for the Chatham Island cluster which displays a unique genotyping profile. Shared BOX profile of North and South Island isolates along with the insignificance of the presence/absence of accessory virulence determinants could be in part facilitated by the dissemination of soil microbes that can occur through a range of mechanisms such as the flight patterns of birds that forage on grass grub and other invertebrate species as well as the application of biocontrol agents. By describing phenotypes of both non-pathogenic and pathogenic *Serratia* spp. and cross-referencing to geographic isolation, it is possible to attribute characterizable differences that, when combined with *in silico* genome analysis (Chapters 5 and 6), could lead to the attribution of traits to chromosomal based evolution. In addition, the attribution of traits specific to a species (DNases) or a pathotype (DNases and siderophore) can begin to allude to the mechanisms these organisms utilise for survival in an ecological niche alongside the acquisition of new beneficial traits. Across all plate assay analyses, the trend for *S. entomophila* (by pathotype) results in a concise dataset of characterised phenotypes with small error bars. This supports the hypothesis that *S. entomophila* is genetically conserved, showing similar production across all isolates in comparison to *S. proteamaculans*, supporting one of the leading hypotheses of this study (Chapter 1, hypothesis i).

Chapter 5

Defining variation in the *Serratia entomophila* genome

5.1 Introduction

As previously discussed, isolates of *Serratia entomophila* and *S. proteamaculans* (Yersiniaceae) are causal agents of amber disease- a host-specific disease of the Scarabaeidae *Costelytra giveni*. *Serratia entomophila* can be considered endemic to New Zealand with only single examples of isolates located outside of Oceania (including France, Argentina, India, and Mexico) (Grimont et al. 1988, Nuñez-Valdez et al. 2008). Previous research of Dodd (2003) and Hurst (Unpublished data) has shown homogeneity in relation to the pADAP plasmid and the chromosome itself in this species. As outlined in Chapter 4, the prevalent disease state of *S. entomophila* isolates is chronic, with a long infection of 3 to 4 months until death. At localities without *C. giveni,* levels of *S. entomophila* in the soil decrease over time. This evidence suggests that *S. entomophila* has formed a host-specific relationship with *C. giveni*, and over time has evolved to cause a chronic disease state. To help understand this, multiple chromosomes of *S. entomophila* pathogens and conspecifics from across New Zealand were sequenced. Genome comparisons would then allow for identification of potential genetic mechanisms that would drive evolution in *S. entomophila*. By identifying regions of genetic diversity or homogeneity, it can be established whether genetic indicators point towards trophic or pathotypic specialisation in *S. entomophila*.

5.2 Results

5.2.1 *Serratia entomophila* isolate 626

Serratia entomophila isolate 626 is currently used as the active agent in the commercially available product Bio-shield®. As the initial 'gold standard' sequencing isolate through PacBio sequencing, *S. entomophila* 626 was used as reference genome PacBio. Additional analyses then extended from the 626 reference to other newly sequenced *S. entomophila* (and later *S. proteamaculans*) isolates generated through the study.

PacBio SMRT sequencing and assembly (Chapter 2, section 2.5.1) of *S. entomophila* isolate 626 generated two contigs, identified as the chromosome and pADAP plasmid. After circularising of the contigs by Circlator to remove overlaps, the chromosomal contig was determined to be 5,027,485 bp in size. Overall, the circularised chromosome of *S. entomophila* 626 was predicted to have a G+C content of 59.1% and encode 4,565 genes. When BLASTn was conducted on the chromosome against the NCBI database the closest (% similarity) sequence was for *S. proteamaculans* isolate 568 (5,448,853 bp). Using Artemis Comparison Tool (ACT) to assess the assembly of isolate 626 against *S. marcescens* (a similar sequence as identified in Chapter 3, section 3.2), the 626 chromosome shares similar genomic synteny to the NCBI reference for *Serratia marcescens* isolate Db11 (Figure 25).

An initial 16S rRNA sequence-based phylogeny of 626 with reference GenBank members of the *Serratia* genus, using the Yersiniaceae *Yersinia entomophaga* as an outgroup, placed *S. entomophila* 626 as closely related to *Serratia marcescens,* with *Serratia ficaria* showing high degrees of similarity to be treated as potentially the same species (Chapter 3, Figure 10). That *S. marcescens* and *S. ficaria* sat more closely to *S. entomophila* in the clade is interesting, as neither of these species has been found to cause disease in grass grub larvae.

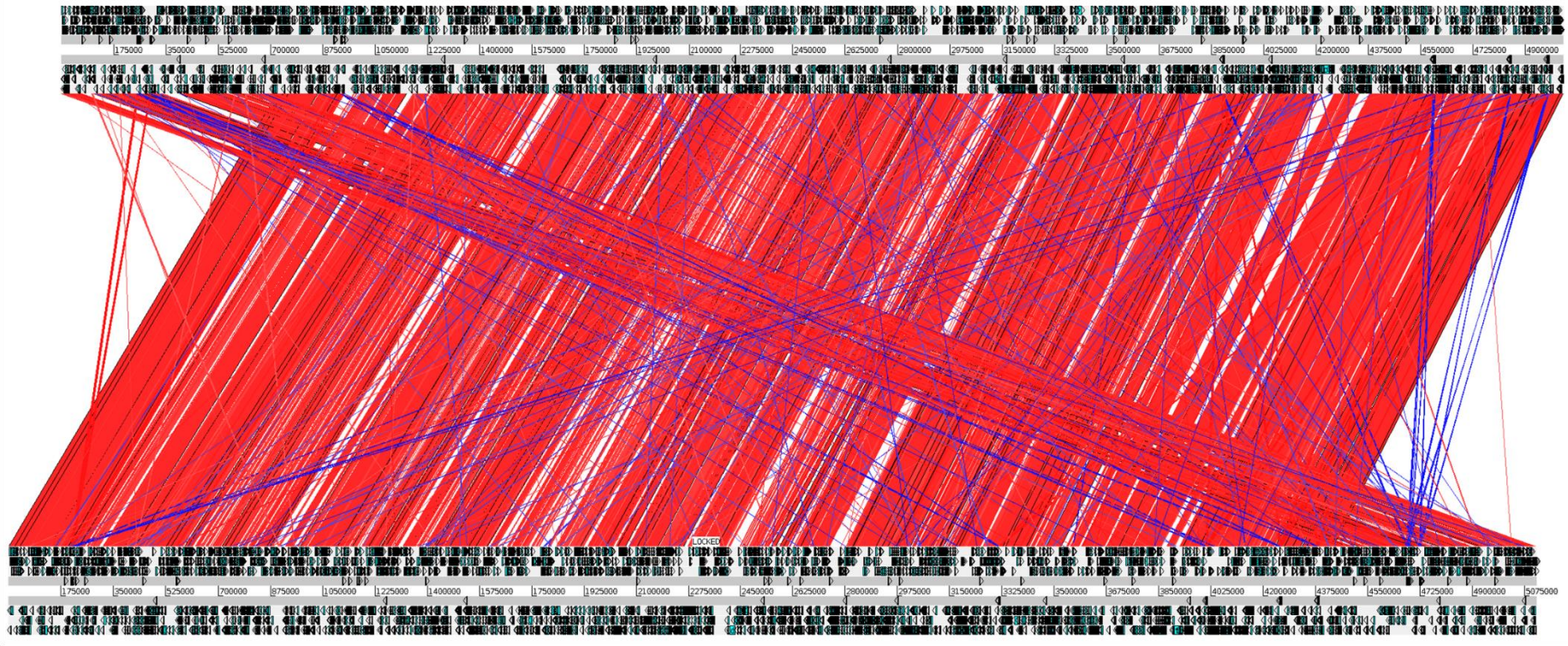


Figure 25 Map of sequence homology from assembly to reference genome
From ACT of *Serratia entomophila* 626 with reference type strain *Serratia marcescens* DB11 aligning chromosomal sequences against one another. Red lines show regions of homology between the two sequences whereas blue lines represent the reverse matches.

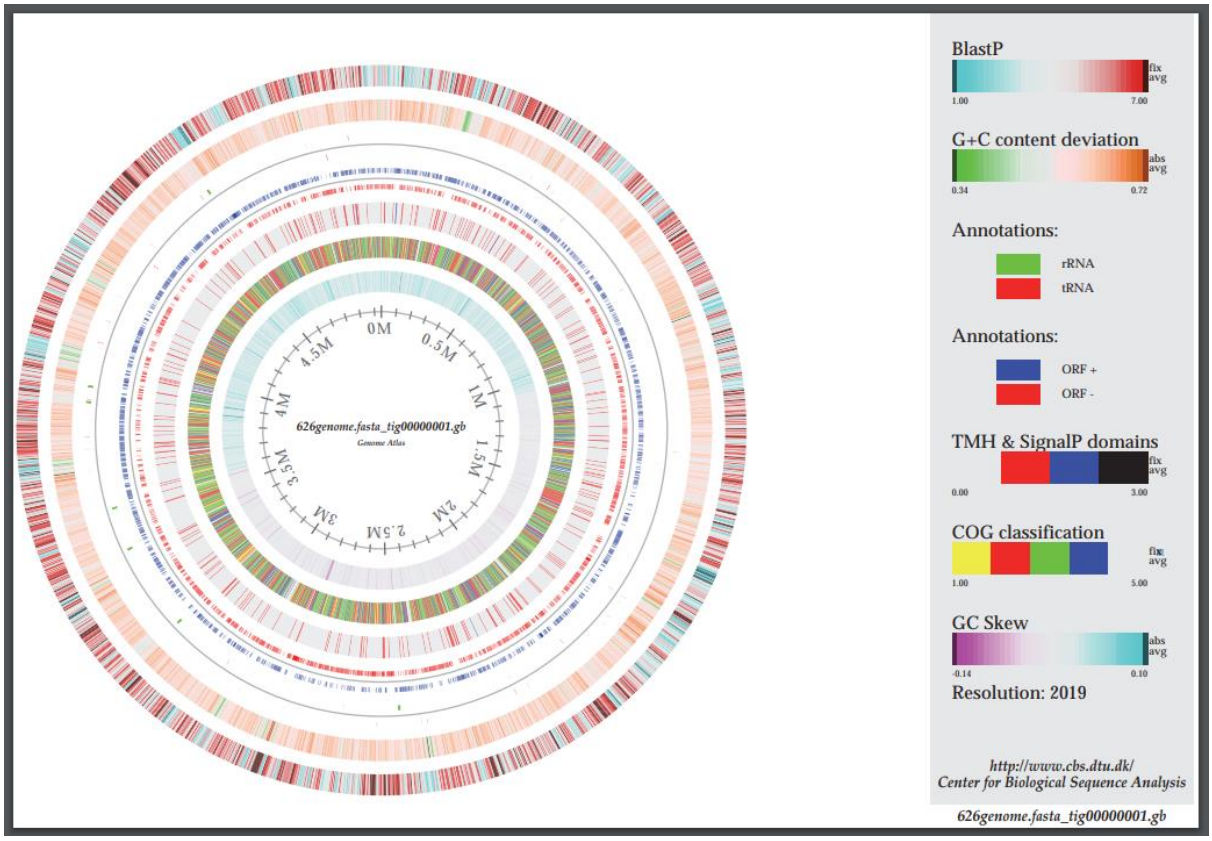


Figure 26 Genome atlas for *Serratia entomophila* 626.
The outermost circle shows BlastP similarities against non-redundant databases. Regions in blue represent unique proteins whereas red indicates high levels of conservation. Circle 2 shows G+C content deviation, where dips below the average G+C content are shown in green, and high spikes in orange. Circle 3 shows annotations of rRNA (Green) and tRNAs encoded on the forward and reverse strand. Circle 4 shows ORF orientation either in sense (+) or antisense (-) orientation. Circle 5 shows predication of Signal peptide domains. Circle 6 shows assigned COG classification assigned into categories 1-5, 1) Information storage processing 2) cellular processes and signalling 3) metabolism 4) poor characterisation 5) uncharacterised or no assignment. The final innermost circle shows G+C skew.

Clusters of Orthologous Groups of proteins (COG) were analysed for the distribution of functional orthologues groups in isolate 626 compared to type strains of other *Serratia* species (Table 12). *S. entomophila* 626 encodes fewer proteins (n=5289) that can be assigned to a COG cluster than *S. proteamaculans* (n=5935). This is consistent with *S. entomophila* isolate 626 having a smaller chromosome than *S. proteamaculans* reference 336X. 626 encodes fewer COGs than any other *Serratia* species, but the percentage genome allocation of these clusters shows key differences. A noticeable difference in proteins assigned to the group relating to energy production and conversion (category C) was found in *S. entomophila*, with an overall lower allocation to this COG category than other isolates. When compared with the *S. proteamaculans* reference 336X (accession NZ_CP045913), there are greater reductions in phage-associated proteins and replication (Category X). *S. entomophila*

626 has more COGs assigned to ribosomal structure and translation (Category J), and carbohydrate transport and metabolism in 626 is approximately 0.6% less than *S. proteamaculans* 336X. As *S. entomophila* sits more closely on the 16S dendrogram (Chapter 3, Figure 10) to *S. marcescens* and *S. ficaria* than *S. proteamaculans*, it might have been considered that the COG classifications would align more with these isolates. This is not the case with carbohydrate transport and metabolism (category G) lower by percentage in *S. entomophila* 626 (7.45%) than either *S. marcescens* DB11 or *S. ficaria* (8%) NBRC 102596 (9%). In addition, *S. entomophila* 626 encodes for more phage-associated genes than either *S. ficaria* or *S. marcescens*. Relatively similar percentages of cell defence associated proteins were found in all species (Table 12). Similar trend can be observed for the COG category Secondary metabolites biosynthesis, transport, and catabolism (Category Q).

As 16S phylogeny placed *S. ficaria* type strain with the *S. entomophila* isolates, comparisons between the two were pertinent. Overall, the number of orthologous groups predicted between the two was similar. However, there are differences in the function allocation of these groups. *Serratia ficaria* NBRC 102596 encodes for more energy-producing (+0.52%) and carbohydrate metabolism (+1.55%) than *S. entomophila* 626 (Category C 4.48% and Category G 7.45% respectively). *S. entomophila* 626 encodes for 0.3% more phage-derived proteins (Category X) in addition to 0.14% more allocation to cell defence mechanisms (Category V) (Table 12). This indicates that there are functional, yet small, differences between both species.

Table 12 COG breakdown of *Serratia entomophila* 626 in comparison to five type isolates of species of *Serratia* from GenBank

The two columns for each isolate assessed show an individual number of COGs labelled as No., then the percentage of the genome made up by this category, listed as %. Total COG's are shown at the bottom of columns.

COG category	Verbose Descriptor	*S. entomophila* 626		*S. proteamaculans* 336X		*S. grimesii* BXF1		*S. marcescens* DB11		*S. ficaria* NBRC 102596		*S. nematodiphila* DZ0503SBS1	
		No.	%	No.	%	No.	%	No.	%	No.	%	No.	%
A	RNA processing and modification	0	0.00	1	0.02	1	0.02	1	0.02	1	0.02	1	0.02
B	Chromatin structure and dynamics	0	0.00	0	0.00	1	0.02	1	0.02	0	0.00	1	0.02
C	Energy production and conversion	237	4.48	275	4.63	260	4.90	253	4.69	259	5.52	248	4.48
D	Cell cycle control, cell division, chromosome partitioning	46	0.87	48	0.81	47	0.89	50	0.93	47	1.00	48	0.87
E	Amino acid transport and metabolism	470	8.89	503	8.48	486	9.15	482	8.94	492	10.49	490	8.85
F	Nucleotide transport and metabolism	104	1.97	112	1.89	116	2.18	101	1.87	108	2.30	104	1.88
G	Carbohydrate transport and metabolism	394	7.45	478	8.05	450	8.47	431	8.00	488	10.41	431	7.78
H	Coenzyme transport and metabolism	222	4.20	218	3.67	223	4.20	229	4.25	242	5.16	245	4.42
I	Lipid transport and metabolism	174	3.29	173	2.91	164	3.09	176	3.27	175	3.73	172	3.11
J	Translation, ribosomal structure, and biogenesis	279	5.28	282	4.75	282	5.31	279	5.18	270	5.75	282	5.09
K	Transcription	433	8.19	489	8.24	478	9.00	477	8.85	460	9.81	484	8.74
L	Replication, recombination, and repair	128	2.42	174	2.93	140	2.64	131	2.43	135	2.88	135	2.44
M	Cell wall/membrane/envelope biogenesis	285	5.39	306	5.16	296	5.57	289	5.36	277	5.91	286	5.16
N	Cell motility	114	2.16	126	2.12	90	1.69	118	2.19	102	2.17	126	2.27
O	Posttranslational modification, protein turnover, chaperones	167	3.16	179	3.02	189	3.56	182	3.38	174	3.71	183	3.30
P	Inorganic ion transport and metabolism	302	5.71	318	5.36	317	5.97	317	5.88	307	6.54	313	5.65
Q	Secondary metabolites biosynthesis, transport, and catabolism	125	2.36	133	2.24	135	2.54	146	2.71	138	2.94	141	2.55
R	General function prediction only	420	7.94	429	7.23	438	8.25	451	8.37	457	9.74	466	8.41
S	Function unknown	255	4.82	279	4.70	253	4.76	273	5.07	260	5.54	280	5.06
T	Signal transduction mechanisms	221	4.18	248	4.18	225	4.24	236	4.38	234	4.99	237	4.28
U	Intracellular trafficking, secretion, and vesicular transport	89	1.68	108	1.82	57	1.07	72	1.34	88	1.87	85	1.53
V	Defence mechanisms	113	2.14	124	2.09	116	2.18	111	2.06	110	2.34	126	2.27
W	Extracellular structures	35	0.66	48	0.81	23	0.43	33	0.61	26	0.55	37	0.67
X	Phage-derived proteins	69	1.30	132	2.22	34	0.64	49	0.91	32	1.68	71	1.28
not defined	Not assigned COG category	607	11.48	752	12.67	489	9.21	501	9.30	457	9.74	547	9.88
Total		5289		5935		5310		5389		5339		5539	

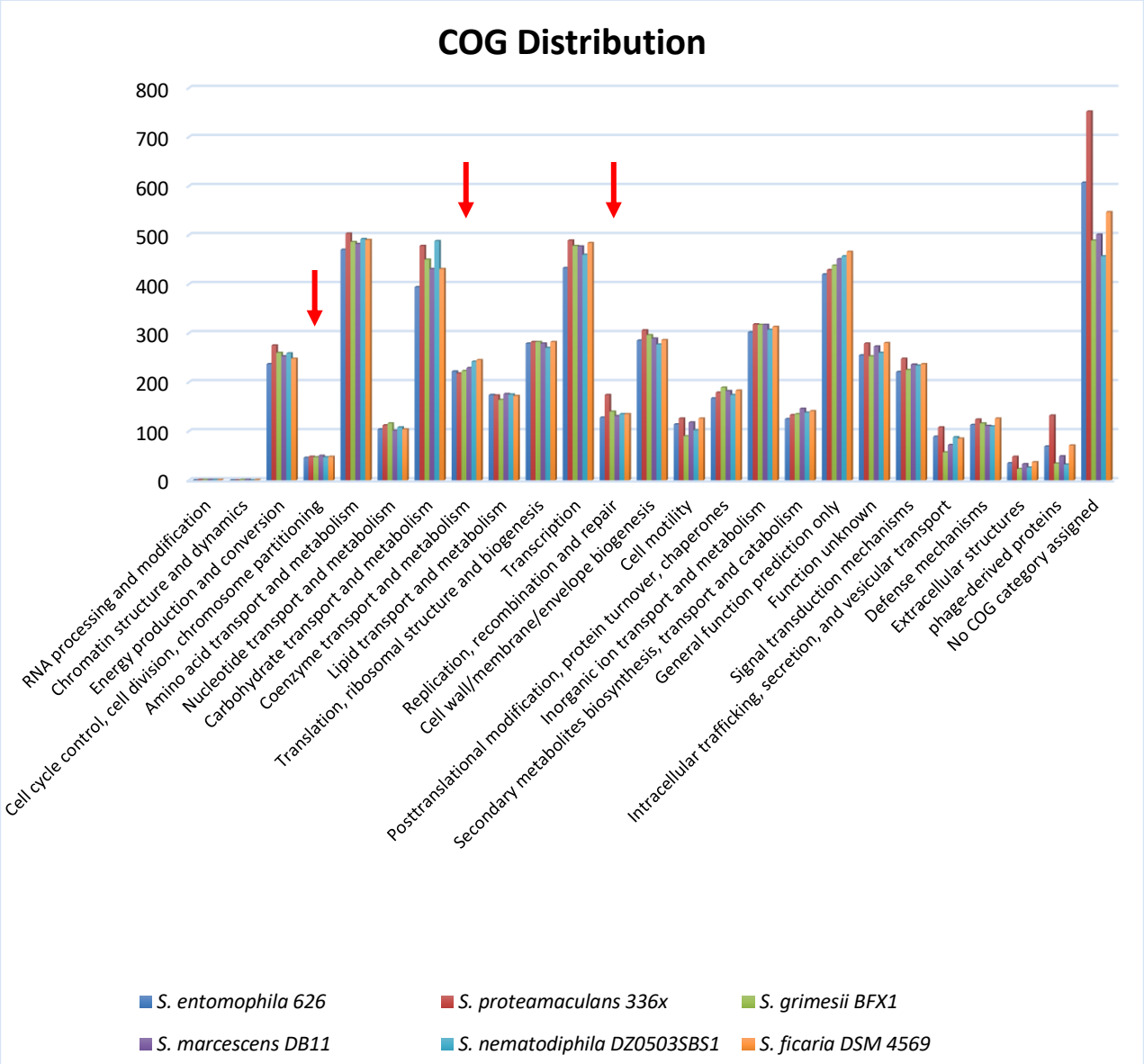


Figure 27 COG distribution comparisons for *Serratia entomophila* and type strains of *Serratia* spp.
Percentage COG distributions of annotated genes and their functions in the complete chromosomes of species belonging to the *Serratia* genus. The cumulative stacked count shown for each species representative.
Number of orthologous groups represented on the Y axis. Full COG breakdowns shown in Table 12.
Red arrow indicates where *Serratia entomophila* 626 shows lower predicted proteins for a COG category.

5.2.2 Genome statistics for *S. entomophila*

Once the 626 *S. entomophila* genome was annotated and characterised, a further 19 isolates of *S. entomophila* were added in addition to one isolate recorded as *S. ficaria* (isolate 457) (Table 13). The inclusion of *S. ficaria* into this analysis was based on 16S phylogenies in Chapter 3. Sequence comparisons of linear *S. entomophila* chromosomes show predictions of tRNAs, rRNAs, and CDS (Table 13). The range of G+C content was from 58.8% (isolate 440) to 59.4% (isolate 364). Over 50% of sequenced *S. entomophila* isolates show an approximate G+C content of 59.2-59.3%, showing limited inter-isolate variation. The number of predicted genes (CDS) ranged between the Chatham's Island isolate 440 (4,765 ORFs) and the Otago isolate 210 (4,373). Pathogenic isolates from Canterbury used as BCAs, *S. entomophila* 626 and A1M02, have similar CDS (4,565 and 4,500 respectively). Isolate 626, which is used as the reference genome in this context, encodes for 86 tRNAs and 10 rRNAs, and has one of the larger *S. entomophila* chromosome sizes at 5,027,485 bp. The Chatham Island isolates have the largest chromosome sizes of any isolate of *S. entomophila*, with isolate 440, 5,200,843 bp, and isolate 442, 5,122,926 bp, which correlates with the larger number of CDS encoded by these isolates (4,765 and 4,702 ORFs). The chromosome size of *S. ficaria* isolate 457 is 5,280,462 bp, 2.54 kb higher than the average *S. entomophila* chromosome. The number of encoded tRNAs and rRNAs of *S. ficaria* isolate 457 fits within the range of those found in *S. entomophila* isolates, though the G+C content is increased at 60%. *Serratia ficaria* 457 has the highest encoded CDS at 4,835. From genome statistics alone, we can see that although there are similarities, the size and G+C content diverge from that seen in *S. entomophila.*

Overall, the characteristics of the chromosomes of each *S. entomophila* isolate are fairly uniform. No G+C content diverges from the average and predicted coding sequences are also in a conserved range. The main source of difference across the chromosomal sequences is the variance of rRNA's encoded, which can be largely split into two groups- 8-13, and >20 (Table 13). rRNAs usually group together as an operon consisting of a 16S, 23S, and 5S rRNA gene. Isolates 626 and 440 both encode 22 rRNA genes that have seven operons, of which six have the usual operon conformity and one has an extra 5S rRNA gene. Six 16S genes from 626 and 440 were all extracted from the genome data and aligned to their gene repeats to determine the level of differentiation. In isolate 626, these six sequences shared 99.5% homogeneity, with few nucleotide differences, and similarly so, the seven 16S genes of 440 share 99.8% homogeneity. The 16S phylogenies from Chapter 3 were tested with several 16S genes from each isolate and showed no change in the overall topology. These isolate groups were assessed for a geographical relationship, as a reduction in rRNA copy numbers has been implicated in rapid adaptation to a change in an ecological niche, potentially increasing a competitive advantage

(Klappenbach et al. 2000). However, no link could be found. Of the distinct geographic groups in this study, the Chatham isolates 440 and 442 were split between these rRNA groups, showing that reduced rRNA encoding operons are not related to a geographical boundary.

Table 13 Summary genome statistics of individual *Serratia entomophila* isolates assessed in the study

	S. entomophila isolate											
Feature	158[1]	176[1]	210[1]	219*	220*	294*	305[1]	345[1]	364*	398[1]	440[2]	Sent1[1]
Size (bp)	4,941,866	4,899,210	4,814,908	4,905,816	4,888,138	5,068,062	5,097,608	4,859,889	5,022,245	4,863,149	5,200,843	5,011,772
G+C content (%)	59.1	59.2	59.3	59.3	59.1	59%	59.0	59.2	59.1	59.4	58.8	59.1
CDS	4,508	4,445	4,373	4,425	4,457	4,648	4,653	4,424	4,558	4,413	4,765	4,597
tRNAs	78	79	77	86	83	83	85	87	83	82	87	88
rRNAs	12	8	12	22	10	9	10	22	13	9	22	22

	S. entomophila isolate											
Feature	442[2]	477*	482*	626[1]	1100[1]	1343*	A1M02[1]	I*	iDIA[1]	MC2*	Man3[1]	**457**
Size (bp)	5,122,926	4,980,089	4,985,006	5,027,485	4,884,015	5,073,106	4,940,707	4,852,108	4,973,092	5,073,553	5,018,275	5,280,462
G+C content (%)	58.9	59.2	59.2	59.1	59.2	59.0	59.2	59.2	59.2	59.0	59.0	60.0
CDS	4,702	4,536	4,632	4,565	4,462	4,651	4,500	4,439	4,516	4,643	4,641	4,835
tRNAs	81	81	81	86	78	84	82	84	90	87	83	84
rRNAs	8	8	10	22	10	10	10	12	23	22	11	8

[1] Isolates encoding pADAP

[2] Isolates encoding a non-pADAP plasmid

* Plasmid free isolates

S. ficaria isolate 457 shown in bold.

ANI values (section 2.5.5) were calculated using sequences of all *S. entomophila* isolates. Values recorded over 95-96% are generally considered to be isolates of a single species ((Kim et al. 2014). All isolates of *S. entomophila* analysed share a common ANI of over 98%, confirming that they represent the same bacterial species (Figure 28). A high percentage of nucleotide identity between the 23 isolates of *S. entomophila* also implies little chromosomal differentiation. The largest measure of distance (<99%) is between Chatham Island isolates 440 and 442, as well as West Coast, isolate MC2 when compared to other isolates of *S. entomophila*. The Chatham Island subgrouping was expected due to its geographic isolation and supported by the differences of CDS encoded as shown in Table 13, but the lower degree of similarity between most isolates and MC2, and the similarity to the Chatham Islands isolates (as discussed in Chapter 4). As Chatham Island isolates and West Coast isolate MC2 are from areas with no grass grub population, this is consistent with other factors (Chapter 4, section 4.2.3) that suggest a potential influence of the grass grub host on the chromosome. However, the lower nucleotide identity between West Coast isolates 219 and MC2 implies that geographic isolation is not the only factor in genomic composition. Although these ANI numbers are very similar, there is still clustering of isolates. No trend between pathotype or plasmid presence was observed. The average ANI between any *S. entomophila* isolate and *S. ficaria* 457 was 91.16%, below the hypothetical species demarcation threshold of 95-96%. This supports the previous statement in Chapter 3 that in similar isolates 16S is not a good delineator of closely related species, with further genes needed to provide accurate phylogeny. The inclusion of *S. ficaria* 457 in the ANI-represented heatmap (Figure 28) would have meant that delineation of the groups of *S. entomophila* was more unclear due to the broader range displayed. For this purpose, the full ANI matrix and heatmap including *S. ficaria* 457 can be found in Appendix C.1.

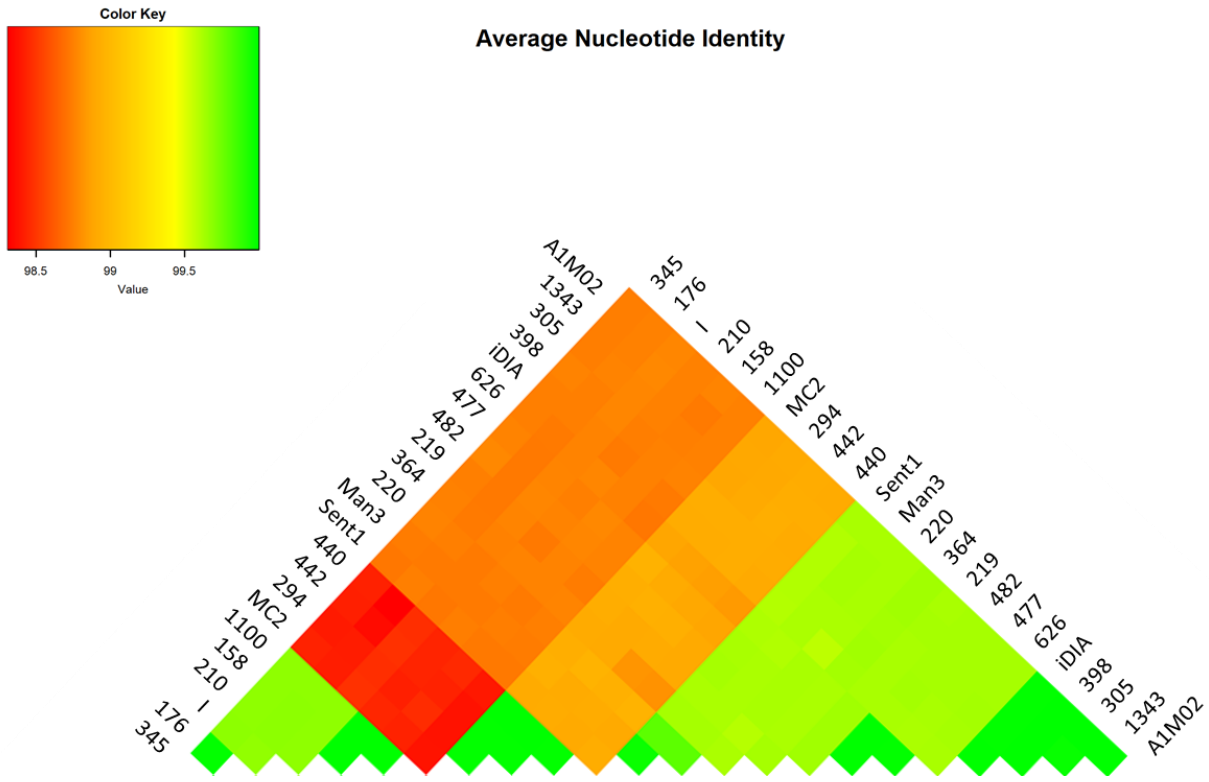


Figure 28 ANI values of all *the Serratia entomophila* isolates displayed in a heat map adapted from matrix from ANI/AAI Matrix.
Green indicates a higher percentage of nucleotide similarity (>99.5%) and red the lowest value (98.4%). Four similarity groups can be seen in the *S. entomophila* isolates, with the main degree of difference observed between Chatham Island (440,442,294) and West Coast (MC2) isolates to various Canterbury isolates.

5.2.3 MLSA phylogeny

As discussed in Chapter 3, the New Zealand *S. entomophila* isolates are so closely related based on 16S sequencing, that a broader range of genes is needed to compare them. Limitations of 16S phylogenies have been previously reported (Poretsky et al. 2014) and demonstrated by the placement of *S. ficaria* within the *S. entomophila* (Chapter 3, Figure 10). A multi-locus sequence analysis (MLSA) was deemed appropriate to further expand on the phylogenetic relationship (Gabriel et al. 2014). For *S. entomophila* and *S. ficaria*, MLSA housekeeping gene data for *rpoAB, recA infB, gyrB,* and *atpD* were extracted from the genome sequences and concatenated for each isolate resulting in multilocus lengths of 12,349-12,570 bp. They were then used to infer a more accurate phylogeny. Here, *S. ficaria* type strain NCTC 12148 was included to assess the delineation of *S. ficaria* isolate 457 to the species type strain (DSM 569 used previously in Chapter 3 was not available as a full genome, only a 16S sequence). As with 16S, the resultant tree rooted with *Y. frederiksenii* clusters *S. entomophila* resolves as between two and four distinct clades with bootstrap support and distinct branch length (Figure

29A). The Chatham Island isolate 442, however, groups with isolate Man3 (Rangitikei) which is not found on the West Coast of New Zealand which Chatham Island isolates have previously clustered with. Chatham Island isolate 294 groups with *S. ficaria* type strain NCTC 12148. This result was puzzling due to the geographic isolation of the Chatham Island isolates and the further similarities between all Chatham Island isolates noted further in section 5.2.4, but further investigation of the MLSA concatenated clustal alignment showed that both these isolates had an insert region in the *infB* gene that was not present in any other sequenced genomes assessed. In addition, the lack of clustering between *S. ficaria* type strain NCTC 12148 and *S. ficaria* isolate 457 shows that again, the delineation between species for *S. ficaria* presented here has not been fully resolved. No further traits related cluster have been identified from the phylogenetic inference as a mix of pathotypes, BOX fingerprinting profile, and location is observed on all branches (Figure 29B). The branching determined from MLSA (Figure 29A) shows degraded clustering that implies distinguishing between these closely related isolates was successful and suggests that *S. entomophila* species is split into four clades. The results from the MLSA support the 16S that *S. ficaria* isolate 457 was closely related to *S. entomophila* isolates, further supported by ANI (section 5.2.2).

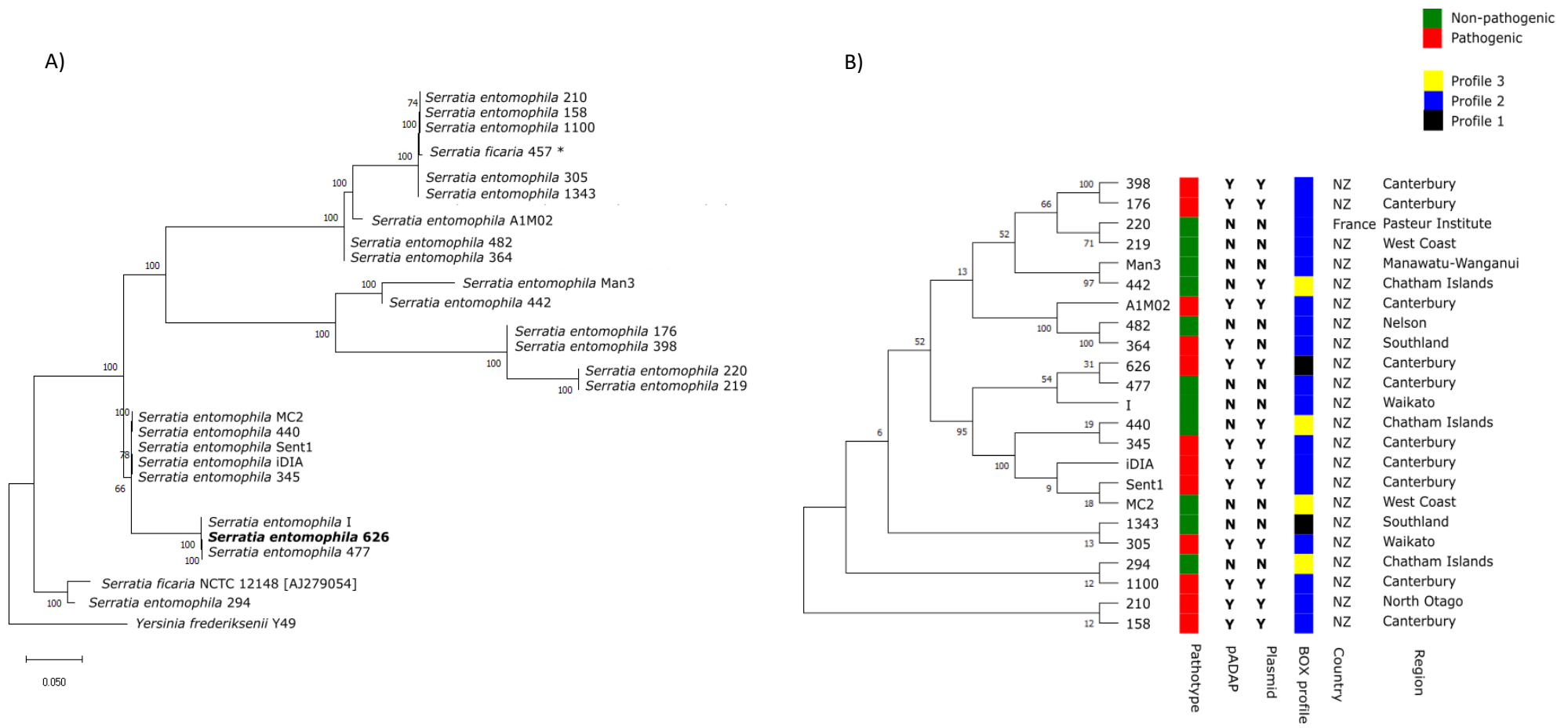


Figure 29 MLSA Phylogenetic analysis

A) MLSA phylogenetic analysis of *Serratia entomophila* isolates with *Serratia ficaria* 457 (indicated with *) and *Serratia ficaria* type strain NCTC 12148 analysed by maximum likelihood. *Yersinia frederiksenii* Y49 used as an out grouping. Evolutionary history inferred by the Tamura-Nei model. This tree is drawn to scale with branch lengths measuring the number of substitutions per site. Evolutionary analyses conducted in MEGA7. B) MLSA inferred topographical tree aligned to known traits and geographic isolation data of *Serratia entomophila* isolates determined in Chapter 4. Pathotype was defined as either non-pathogenic i.e., showing no disease state, or pathogenic, encompassing all levels of chronic disease. Profiles 1, 2 and 3 refer to BOX profile assessed in Chapter 4, section 4.2.3.

5.2.4 Genome-wide analysis

Using these 23 sequenced *S. entomophila* isolates and the *S. ficaria* isolate 457, the pan and core genomes were identified and analysed for consistency, divergence, and differentiation between each isolate and any traits. In this study, 104,444 putative genes were identified across the pangenome, with 7131 putative genes comprising the orthologous core (Figure 30).

As identified by Roary there are 3878 CDS defined as 'core' in the *S. entomophila* chromosome across all 23 isolates, with genomes having as few as 0 and as many as 125 unique genes in comparison to other isolates. Isolates I, 364, 158, and Man3 have the largest number of unique genes, although this characteristic isn't expected to relate to pathogenicity as Man3 is defined as generally non-pathogenic, with indicators of subtle transient virulence. iDIA, while a chronic pathogen, does have a faster infection rate than isolate 626 and encodes for 60 unique genes. The rapid onset of disease could potentially be mediated by some of these unique genes. By manipulating the .csv output file of Roary to look solely at the genes only associated with iDIA, there was no known association between unique genes to iDIA and pathogenicity (Appendix D.2). There are six additional transposases (five family IS4 and one family IS3), where the presence of these genes has potential implications for horizontal gene transfer. ISfinder was used to validate the presence of these transposases genes on the chromosome of iDIA where IS elements are described as having flanking transposases up and downstream of the insertion. These insertion sequences are usually small containing one open reading frame encoding transposases. Seven transposases (IS4/IS3) were found on iDIA with a single open reading frame with no matches when compared to *S. entomophila* 626.

Visualising the core genome alignment highlights regions of similarity between similar isolates (Figure 31), where isolates can then be divided into sub-groups of *S. entomophila* isolates for comparison. Using this approach, isolates can then be investigated for any beneficial traits for the organism that are unique to specific isolates or pathotypes. The addition of *S. ficaria* isolate 457 shows the variation in the core and accessory genome that is more dissimilar to *S. entomophila* (with orthologous proteins defined as those above 95% similarity using the Roary algorithm). Using the core genome analysis (Figure 31) of the assessed isolates, MC2 was found to be more closely related to 440 and 442. Another cluster of *S. entomophila* isolates that share a larger region of their genome (1100, 210, 158, 176, 345, I) was observed (Figure 31). None of these isolates were from the same region (Figure 29B), with distribution from Hawkes Bay in the North Island to Otago in the South Island. However, like other *S. entomophila* isolates, these are all chronic pathogens.

The Scoary output toolkit defined in Chapter 2, section 2.5.8 was used to determine if there were any clustered genes of interest in horizontally acquired regions (as virulence factors are known to cluster in genomic islands), or if there was evidence of genome reduction in non-pathogenic isolates, and the full results of this analysis are provided in Appendix C.3. Using ten permutations against pathotype, plasmid presence, and the group of Chatham Island isolates, 642 genes returned a naïve P-value of <0.05. Of these 642 potentially associated genes, none were 100% unique to pathogenic isolates. The highest degree of association in *S. entomophila* pathogenic isolates was four genes also encoded for in five of the non-pathogenic isolates (Table 14). When assessing for Chatham Island-associated genes, 810 genes were listed with a naive P-value of <0.05. Of these, one was associated with all isolates except the Chatham Islands, while 23 were solely associated with the Chatham Islands. With the addition of the West Coast isolate MC2, a further 148 genes were found outside of these isolates with 443 addition genes only found in the Chatham Island isolates and West Coast isolate MC2.

Table 14 BLASTp of the highest associated genes with putative pathogenicity traits using Scoary.

Gene	BLASTp	Similarity	Accession
	D-serine/D-alanine/glycine transporter	98.7% *Serratia ficaria*	A0A240BTK8
	Uncharacterised protein	96.2% *Serratia nematodiphila*	A0A086GBE2
	DNA-binding response regulator	83.3% *Serratia liquefaciens*	A0A515CRJ5
	Outer membrane protein TolC	91.9% *Serratia ficaria*	A0A240BPA3

Chromosomal genes encoded are extremely similar between these four isolates showing large, conserved blocks not found in any other chromosome (Figure 31). These three isolates encode two proteases not found on any other *S. entomophila* (*clpP2* and *ydeA*), although lack an encoded extracellular protease found in the 17 others. Six genes showed an association between traits by the Scoary algorithm for plasmid-bearing isolates. Of these, none were unique to plasmid-bearing isolates, therefore we can eliminate them in relation to identifying a unique trait associated gene.

Whole-genome comparisons again support the ANI values described previously that *S. entomophila* and *S. ficaria* are separate species Roary whole-genome alignments have shown the dissimilarity present in comparison to *S. entomophila*.

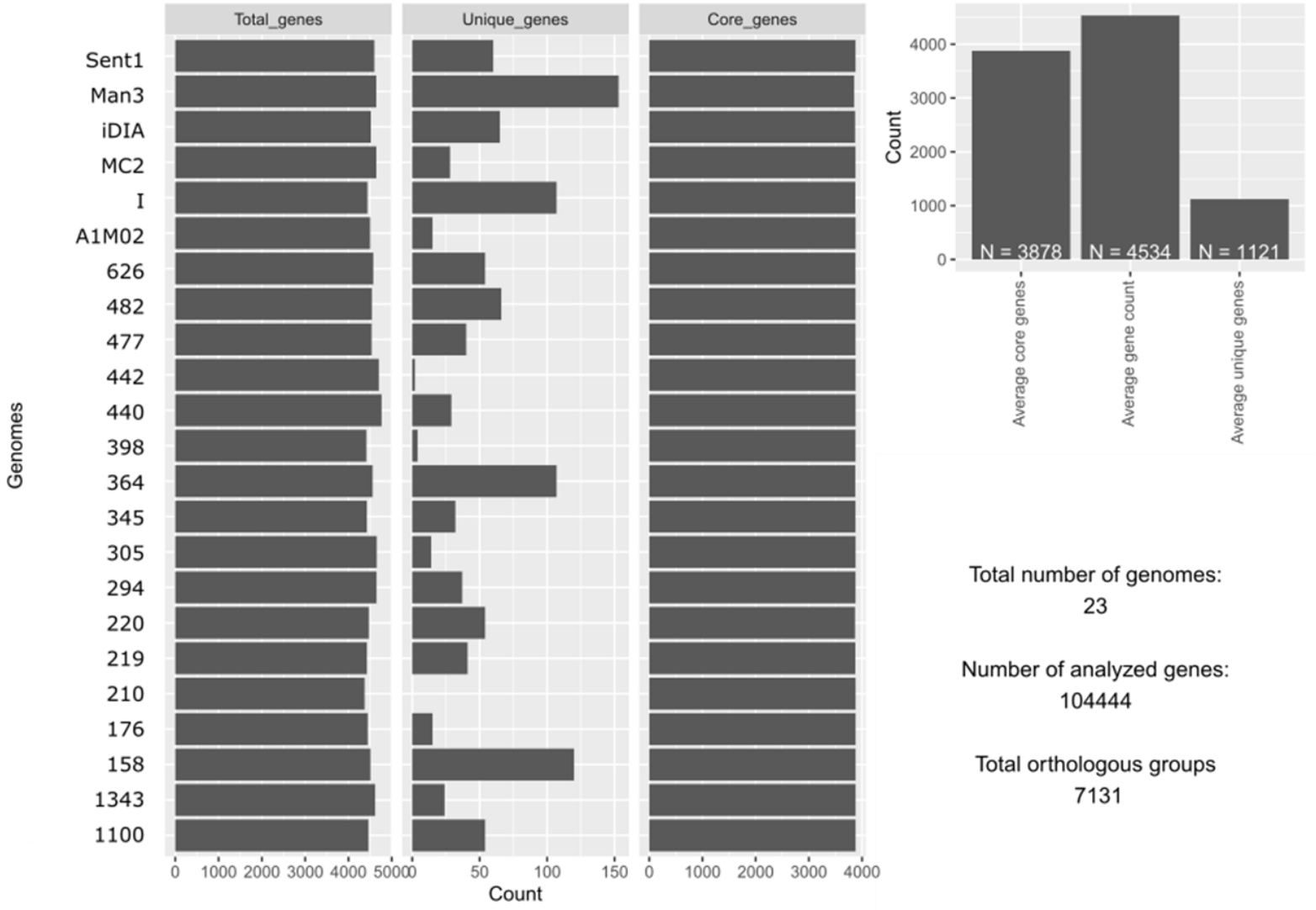


Figure 30 Comparative genome statistics of the core genome of *Serratia entomophila* defined by Roary

Core genes indicates number of genes found uniformly across the annotations of all inputted sequences. Statistics with the addition of *Serratia ficaria* 457 isolate can be found in appendix C.4. Refer to Figure 31 for further detail.

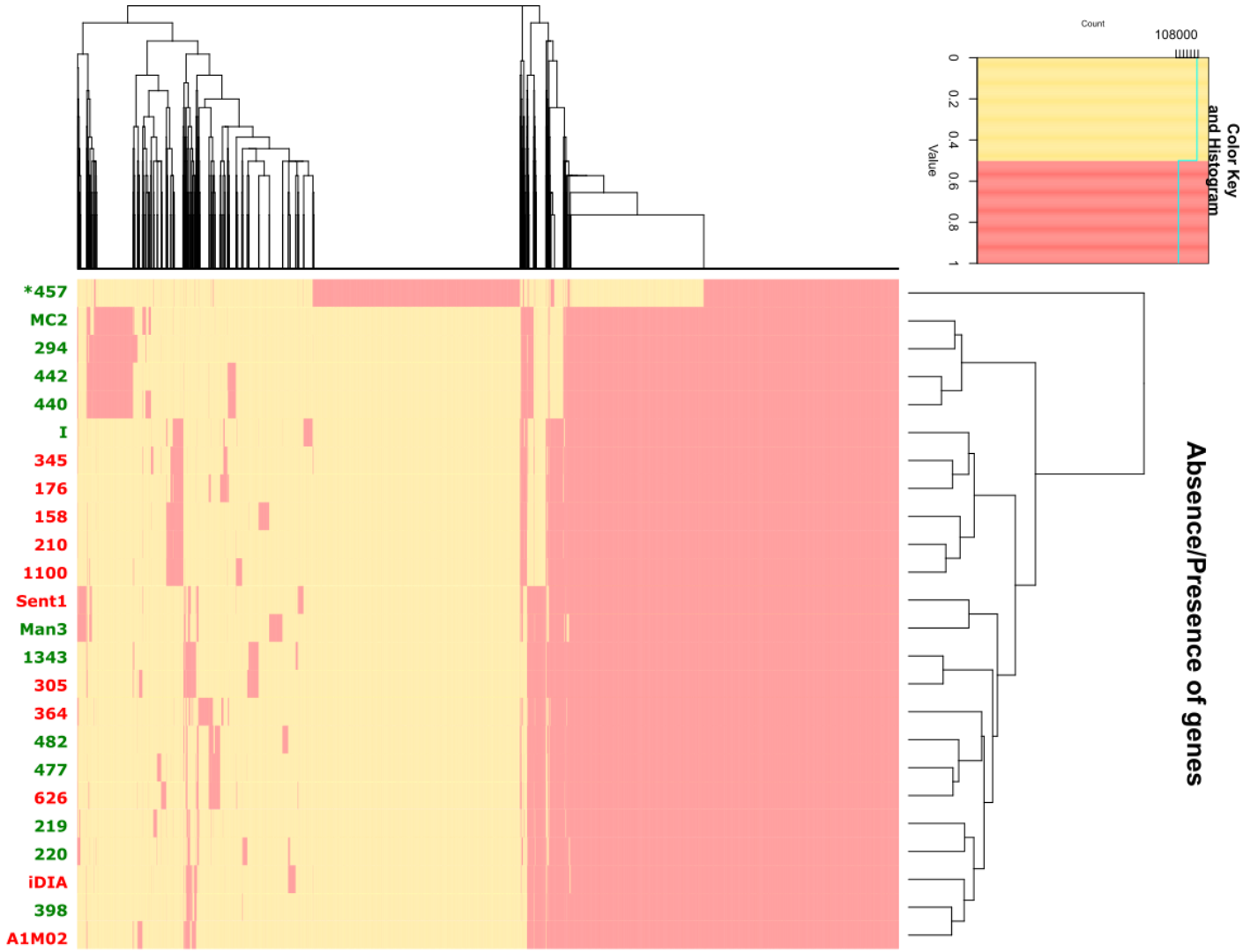


Figure 31 Core genome alignment of all chromosomes of *Serratia entomophila* and *Serratia ficaria*

* Indicates *Serratia ficaria*

Pathogenic isolates highlighted in red isolate number, non-pathogenic in green. Image showing just *Serratia entomophila* can be found in Appendix C.5. Presence/ absence matrix used to generate this data can be found in Appendix C.2.

5.2.5 The openness of the *S. entomophila* genome

The openness of a pangenome indicates whether a species has an upper limit to the number of genes that can be incorporated as new isolates are added to the analysis. No upper limit to the number of genes with each included isolate would define the pangenome as being open, whereas a maximum capacity would be reached in a pangenome that could be described as closed (Park et al. 2019). For diverse genomes, it would be expected that with an increasing number of isolate chromosomes added to the analysis, the overall total of conserved genes would decrease, and the total number of genes would increase, as depicted in Figure 32. However, in the current study after the addition of the second genome to the analysis, the number of conserved genes remains steady and remained steady even with the increase of total genes when n=23 genomes. The implication from this is that the core chromosome of *S. entomophila* is mostly uniform in its genetic makeup.

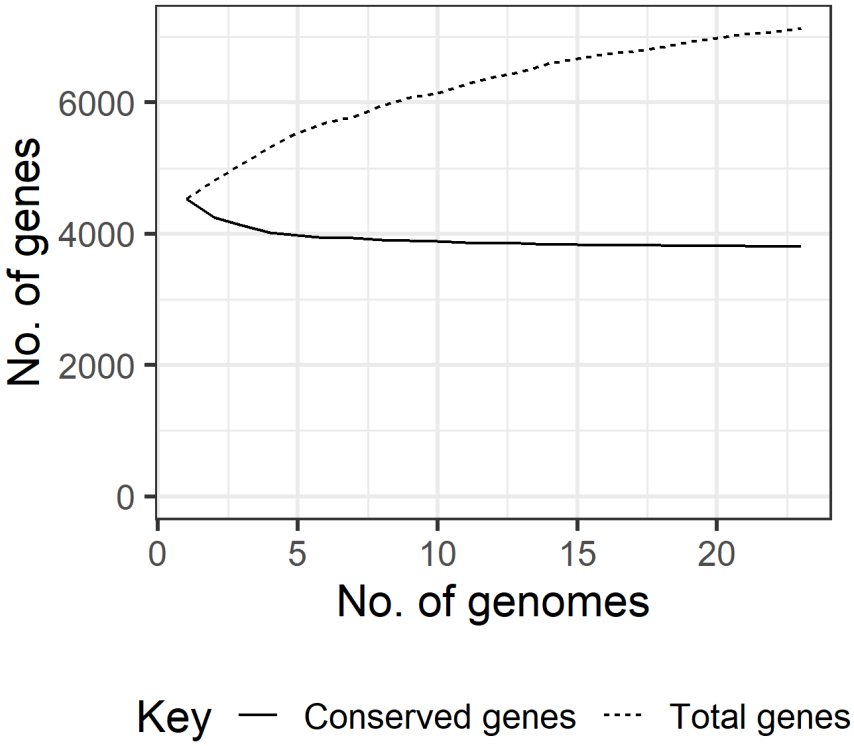
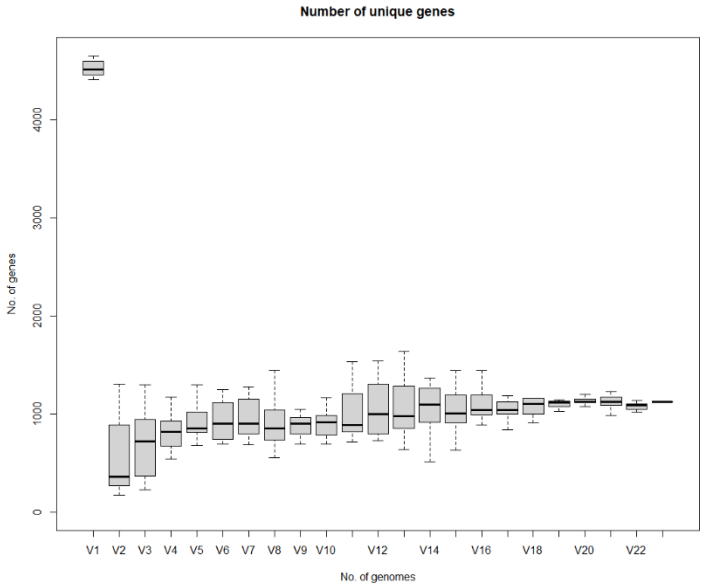


Figure 32 Core genome reduction analysis of sequenced *Serratia entomophila* chromosomes from the core genome compiled with Roary pangenome software.

For each genome added to the dataset, the number of unique gene clusters increases (Figure 33A). The steady increase of total genes in the pangenome with further chromosome additions (Figure 33B) implies that, at this level, the genome is open. To further define, the data was fitted to Heap's Law, which measures the diversity of a pangenome by calculating an alpha number. Values under one

predict a variable pangenome which is described as open, where each genome has many unique genes.

A)



B)

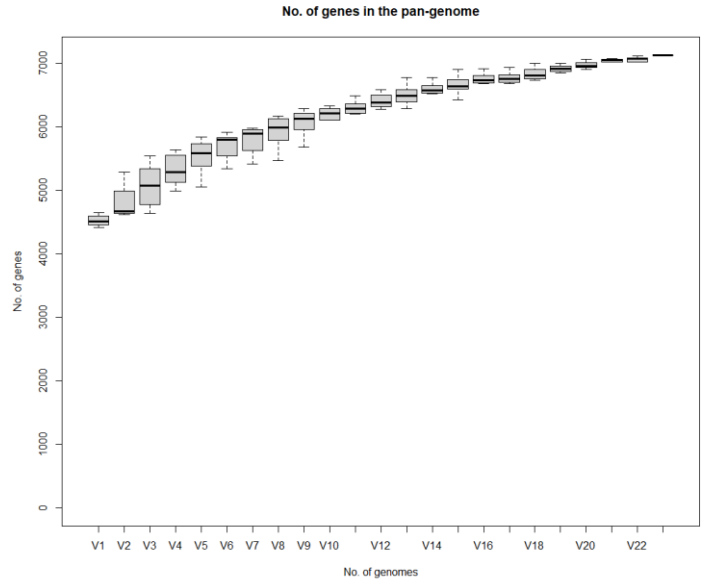


Figure 33 Pangenome analysis made in pangenome plots utilising Roary output.
A) shows the number of unique genes with the subsequent addition of further genes (with the most unique being the primary genome) B) shows the Rarefaction curve of *S. entomophila* pangenome.

Using the R package micropan (Chapter 2, section 2.5.9) the alpha value of the *S. entomophila* pangenome was calculated under controls of 1000 permutations as 0.934, meaning the population size is unbounded. Undertaking an analysis that predicts the lower limit of the pangenome size (Chao estimate) returned an estimate of approximately 6841 unique gene clusters. This is generally a conservative estimate and lies approximately where the model predicts the rarefaction curve would flatten if infinitely more genomes were included (Figure 33B). Consequently, this analysis would be

more useful with a larger number of genomes. With 23 chromosomes of *S. entomophila* isolates from varying geography, pathogenicity, and plasmid bearing capabilities, the core genome remains the same but with each additional genome, unique genes are still added. What can be assumed from the compiled data is that although the model predicts that *S. entomophila* pangenome is unbounded, gene acquisition occurs slowly and to a limiting degree. This result does not support the hypothesis that the *S. entomophila* genome is reducing and becoming more specialised with its host, showing that the opposite is true. Symbionts tend to have a smaller closed pangenome as opposed to those community species with high rates of gene transfer. As a soil-dwelling bacterium, *S. entomophila* encounters other microbes frequently.

Orthovenn2, an interface platform for comparative analysis of orthologous gene cluster, was used to display the number of shared predicted proteins between a representative sequence from each clade grouping. This was used as a reference for how many shared proteins there are between each group, which gave a numeric in addition to the image display of Roary (Figure 31). Based on Figure 31, the isolates 440, 626, 1100, and A1M02 were selected. Most genomes across all isolates share a 'main core'. The largest drop-off in shared genes occurs when sequences MC2, 440, and 442 are separated, as would be expected from Roary pangenome analysis.

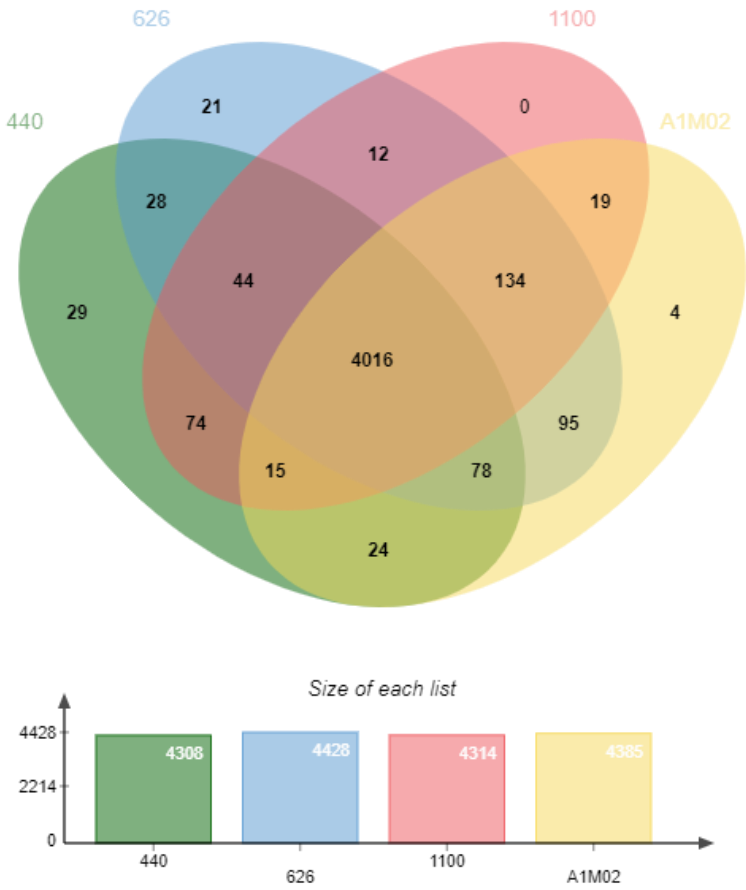


Figure 34 OrthoVenn representation of predicted proteins shared between representatives of each monophyletic group.
Representation of the orthologous clusters, with complete clusters per isolate found in the lower graph. Isolates were selected based on the core phylogeny generated by Roary to encompass each phyletic group.

5.2.6 *Breseq*

Since all isolates of *S. entomophila* were shown to be closely related, and *S. entomophila* isolate 626 was well-annotated based on a PacBio sequence, *Breseq* was used to search for small differences that might elucidate any potential differences between pathogenic reference isolate 626 and non-pathogenic *S. entomophila* isolates. Raw Illumina reads of *S. entomophila* isolates 477 and 440 were independently mapped to the reference sequence to find any consistent differences that could alter coding regions to a degree that would affect the bacterium.

Initial analyses focused on 626 and the non-pathogenic isolate 477, due to their genomic similarity. In total, 98.7% of reads from the forward (16,910,059 reads) and reverse (16,908,861 reads) read files of 477 mapped to the reference sequence. One hundred percent of these reads passed *Breseq*s error

filters to be mapped to 626. Overall, 180 mutation predictions were listed in addition to two unassigned coverage regions. Initial searches of the predicted gene descriptions were undertaken for predicted mutations in regions of interest, i.e., accessory virulence, toxins, or plasmid-related genes. Using *Breseq* no gene interruption of plasmid-associated genes was found between reference 626 and query of 477 that could impact plasmid bearing/ non-plasmid bearing ability.

Breseq assessment was undertaken between 626 and Chatham Island isolate 440, representing the most significantly different chromosomes and being genomes derived from grass grub populated and non-populated regions. In total, 83.5% of reads (14,496,266) were mapped to isolate 626. Thirty-seven SNPs and small indels were assigned to a 440 *paiB* gene- a protease synthase protein involved in the reduction of extracellular protease levels. Ten SNPs were also located in the 440 *lon1*, encoding an ATP-dependent serine protease. The resulting genome searches confirmed that this gene is not predicted by annotation software used for the 440 chromosome. Other protease encoding genes include *ptrB* and an extracellular serine protease. Arguably, one of the most interesting differences shown was that several SNPs resulted in the replacement of a *ptrA* (protease 3) in 440 with coenzyme PQQ synthesis protein that is also found in *S. ficaria*. This *ptrA* gene encodes a seralysin that is implicated in the inhibition of insect antibacterial peptides, notably in the haemolymph by *Galleria mellonella* (up to 50% reduction of synthesis) (Cabral et al. 2004). BLASTn analysis of this gene across all isolates of *S. entomophila* found the change to be uniform across all the assessed isolates. All isolates encode a single copy of the *ptrA* gene, with isolate 626 having two copies. The addition of this extra protease copy implicated in haemolymph seralysin production previously, in *S. entomophila* isolate 626 could be beneficial for survival in the host haemolymph in the final stages of amber disease.

Overall, due to the similarity required for *Breseq* to be wholly effective (this program works by finding mutations in relation to a reference sequence), it was not suitable to be applied to the whole dataset. At 99.99% DNA identity to isolate 626, 477 still identified 180 predicted mutations, with isolate 440 predicting sequencing miscalls and thousands of small genomic changes to the reference. Due to this, comparisons for more varied chromosomes than between isolates 477 and 626 were compared using Mauve alignments (section 5.2.7) and HMM searches for motifs of interest to this study (section 5.2.10).

5.2.7 Mauve

Mauve alignments were constructed to observe any colinear changes visually and statistically between *S. entomophila* isolates. The ability to visualise genome rearrangements and inversions aids

characterisation of these changes across the breadth of isolates assessed in this study. Based on preliminary findings from Roary and phenotypic similarities (section 5.2.4), the initial comparison used was between isolates 626 and 477 (Figure 35). These isolates are both sampled from the same geographic region (Canterbury) and observed to be similar in core genome structure (Figure 31) and 16S phylogeny (Chapter 3, section 3.2). Genomic rearrangements can be seen in two collinear blocks (shown in green and purple). Much of the chromosome (approximately 4.5 Mb), is uniform in its structure on alignment. Isolate 477 has a large 33 Kb region located at approximately 3.1 Mb on the chromosome as drawn that is not found in 626, and conversely, 626 has a similarly sized region at chromosomal position 4.4 Mb that is not found in isolate 477. Full annotation of both regions can be found in Appendix C.6 and C.7. This ~40 Kb cluster of predominantly hypothetical proteins in isolate 626 is similar to the unassigned missing coverage region found in *Breseq* and encodes a predicted phage tail protein (GpFI) and tyrosine recombinase (*xerD*). When the amino acid sequence of GpFI is inputted into UniProt, 82.5% identity is returned to a phage tail monomer from *Ewingella americana* (ATCC 33852) which suggests similarity but is too low to imply homology. Flanked 5' by a tRNA (Met), the potential exists for this region to have been horizontally acquired. This region was identified as a putative genomic island, as described in section 5.2.9. A lysozyme (*rrrD*) and a restriction-modification (R-M) system are located in this cluster, the latter comprising of a type II restriction enzyme ApII 5' of modification methylase paeR71M. On further investigation of this *gpFI* by undertaking BLASTn on all *S. entomophila* isolates five gene hits were matched from 626, 442, MC2, 440, and 1100. All these isolates have the full ~40 Kb region (indicated in Figure 35 with a black arrow on isolate 626) with a flanked tRNA at the 5' end and *xerD* gene at the 3'. Nucleotide identity is high between these isolates. Both Chatham Island isolates (440, 442) and the West Coast isolate MC2 have 85.7% nucleotide identity to this region in isolate 626, while 1100 has 90.3%.

When exploring the additional 33 Kb region found in isolate 477 (Appendix C.7) 36 predicted genes are identified of which 33 are annotated are hypotheticals. A *dam* methylase gene was identified alongside *cysH*, a phosphoadenosine phosphosulphate reductase, and a *xerD* tyrosine recombinase. *cysH* is a second copy of a gene operon found in all isolates of *S. entomophila*. BLASTn of the hypothetical proteins in the operon 3' of the dam/*cysH* region revealed low percent similarities to the hydrolase family protein from *S. liquefaciens* (27.6%), alongside a phage portal protein (70.2% similarity) and a terminase (63.3% similarity). In summary, based on these predictions, the DNA in this region is phage derived, although differing significantly from the closest hits using BLASTn.

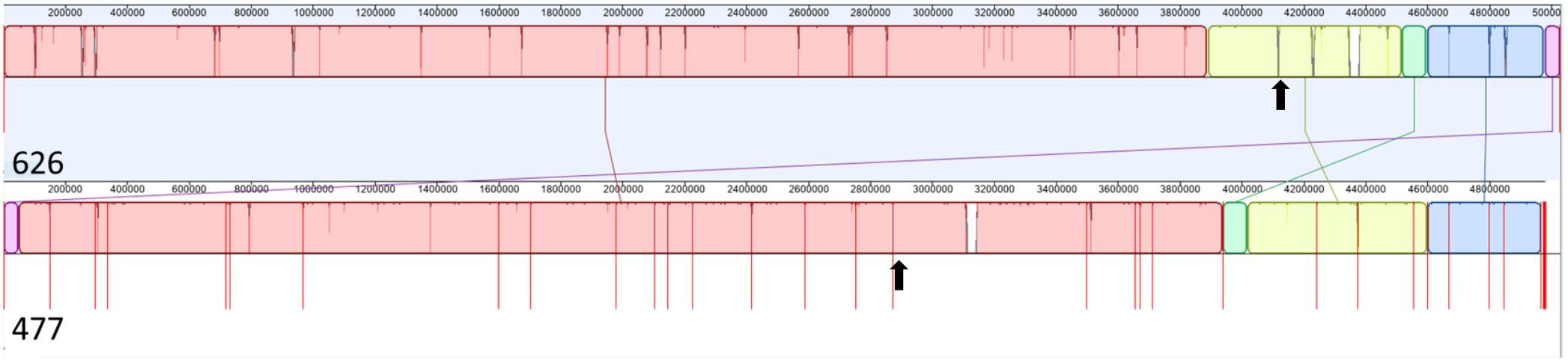


Figure 35 Mauve alignment of *Serratia entomophila* isolates 626 (top) and 477 (bottom)
Comparison to show differences between two extremely similar isolates (by ANI, core genome and phylogeny) of different pathotypes to demonstrate chromosomal variation between closely related isolates. Isolate 626 is a chronic pathogen whereas 477 shows no virulence and has no plasmid. Redlines indicate the contig boundaries of 477 illumina sequence when aligned to isolate 626. Collinear blocks shown in corresponding colours and connected via lines to show regions between the two genomes. Arrows indicate the 40kb region in 626 discussed in is section 5.2.7 and identified using Breseq and the ~40Kb region in 477.

Figure 36 Mauve alignment of multiple sequenced isolates.
Comparison of various isolates of geographic isolation and pathotype of *Serratia entomophila*. (Top- bottom) 345, 626,219, MC2, 440
345 is a multi plasmid bearing isolate found in Canterbury, whereas 219 and MC2 are both plasmid free isolates from the West Coast of the South Island. 440 is a Chatham Island non-pathogenic isolate that is similar in chromosomal alignments, to West Coast isolate MC2 though which bear a non-pADAP plasmid.

Further Mauve alignments were undertaken to visualise the differences between chromosomes of *S. entomophila* of differing pathotypes or regional diversity (Figure 36). The Chatham Islands isolate 440 was selected as a potential outlier to show the difference between these more heterogeneous isolates. . The chromosome of both 626 and 440 share three collinear blocks with varying degrees of homology. Isolate 440 has two large regions on the chromosome which are not present in the 626 chromosome. The first region of differentiation on the chromosome of isolate 440 is located around 2.07 Mb-2.15 Mb as a 35 gene region flanked by tRNAs. This region encodes for two plasmid partition proteins (*xerD*) as well as a second chromosomal *lexA* repressor and a prophage integrase protein. A lysozyme (*rrrD*) was also found, likewise encoded in the prior discussed 40 Kb region from isolate 626, indicating some shared genes. Overall, most of the genes encoded in this region are annotated as hypothetical proteins. This region has a much lower G+C content (50.9%) than the remaining chromosome (~59%) and is absent in isolate 626. Interestingly 626 has none of the genes from this region observed in 440, indicating that this additional 2 Mb region has been either acquired via HGT after the Chatham Island isolates separated from other *S. entomophila* isolates during evolution, or lost after coevolution with grass grub.

Unlike the West Coast isolate 219 the West Coast isolate MC2 was found to be relatively similar in its chromosomal sequence to the Chatham Island isolates. Mauve was used to discern any dissimilarities between MC2 and the Chatham Island isolates. A large region of 440 (~90 Kb) was absent in the MC2 chromosome at position 1.15 Mb in addition to the previously described region missing from the broader isolate sample set (Figure 37). Encoded in this region are genes associated with gene acquisition i.e. transposases. The ClpV1 protein, a component of type six secretion systems (T6SS), was also identified. With the identification of ClpV1, BLASTp was performed on surrounding regions to elucidate whether a T6SS was encoded. The extended region was then run through SecReT6 to predict potential T6SS genes (Li et al. 2015), which identified a putative secretion system in this region, including a type VI effector protein. Similar to the 40 Kb region previously discussed, the 90 Kb region encoded an abundance of hypothetical proteins. BLASTn comparison of this region was used to discern whether there were any DNA similarities to other organisms. Results of BLASTn show a high percentage identity of 94.27% (but only 54% coverage), to a region found on *S. proteamaculans,*

isolate 568, in addition to 93% similarity to *Y. pestis* (49% coverage) and strains of *S. ficaria* and *S. marcescens*..

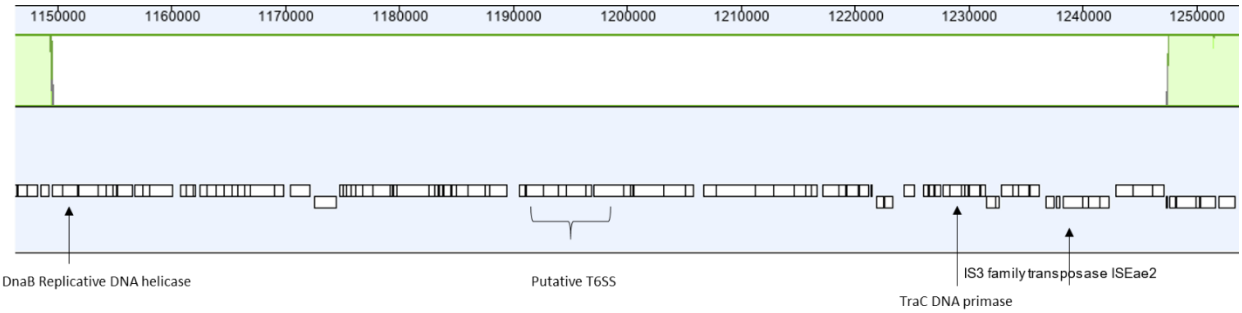


Figure 37 Large ~90 Kb region in Chatham Island isolate 440 that is not present in West Coast isolate MC2.
The 90 Kb region distinguishes Chatham Island isolates from West Coast isolate MC2, which are otherwise highly similar. Green regions at 1.15 Mb sand ~1.25 Mb indicate the regions of homogeny between 440 and MC2, with the white open region being unique to isolate 440. Arrows indicate annotated genes predicted by Prokka software. The bracket shows ORFs predicted as a T6SS. Most genes in this region are listed as 'hypothetical'.

To show an overall comparison of *S. ficaria* isolate 457 to *S. entomophila*, a mauve alignment was constructed mapping the 457 contigs to *S. entomophila* isolate 626 and 440(Chatham Islands) (Figure 38). Alignments were started at the chromosomal replication gene *dnaA*. Although there are large blocks of conservation observed, the similarity plots show divergence between these regions and large gaps between the three isolates. Whereas chromosomal similarity can largely be seen between the arrangements of 626 and 440*, S. ficaria* 457 shows large areas of inversion and relocation on the chromosome. For this study, based on this and previous results it can be assumed they are separate species. As *S. ficaria* isolate 457 is an overseas isolate (France) with no pathogenic effect on *C. giveni* larvae, isolate 457 will no longer be included in further comparisons in specific genes.

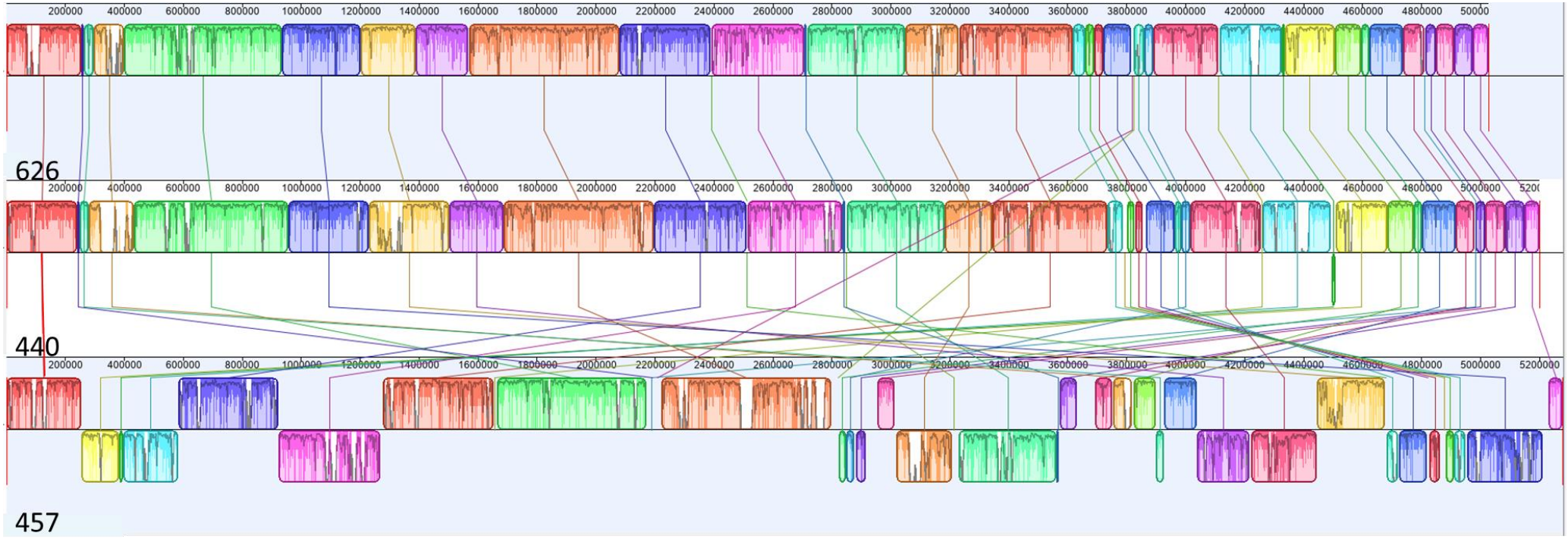


Figure 38 Mauve alignment of isolates *Serratia entomophila* 626 (top) and 440, and *Serratia ficaria* 457 (bottom)
Comparison to show differences between two extremely similar isolates (by ANI, core genome and phylogeny) of different pathotypes to demonstrate chromosomal variation between closely related isolates. Isolates 626 and 440 selected on the basis of 626 being the 'gold standard' reference *Serratia entomophila* sequence. Chatham island 440 as one of the largest *Serratia entomophila* isolates was used to to see any similarities with the similarily sized *Serratia ficaria* genome.

5.2.8 Targeted mutagenesis of a phage-like cluster on the *S. entomophila* chromosome

A 17,224 bp phage-like region flanked by tRNA's was identified through comparisons of the *S. entomophila* and *S. proteamaculans* chromosomes and genomic island predictions (Chapter 5, section 2.5.12). This region encodes a DNA damage-inducible protein I (*dinI*) gene. *DinI* is documented in mediating the SOS response in strains of *S. marcescens (*Berkmen and Benedik 2002*)*. Flanked by tRNA's and a differing G+C to the main chromosome, the region is likely to have been acquired by HGT and could be useful for determining the evolution of *S. entomophila*. The region encodes several phage genes and likely encodes a mature phage-like particle with no phage head. Due to the absence of DNA packaging-associated genes, it is not likely to harbor DNA or to replicate, raising the possibility that it may instead transport the DinI protein. Phage tail-associated genes are present alongside the scaffold, but no other phage constituents are present in the 17.2 Kb region. This gene was not flanked by a phage carrier in any strains of *S. proteamaculans* and was unique to *S. entomophila.* Custom BLASTn from Geneious Prime software was used to identify two *dinI* genes across *S. entomophila* isolates, whereas *S. proteamaculans* only encoded one. All *S. entomophila* isolates were found to encode this phage-like region. Based on the presence of this region in *S .entomophila,* but not the later discussed *S. proteamaculans* and the possible role of Din in potential SOS response repression of *S. entomophila* mediated virulence, the region was subjected to targeted mutagenesis.

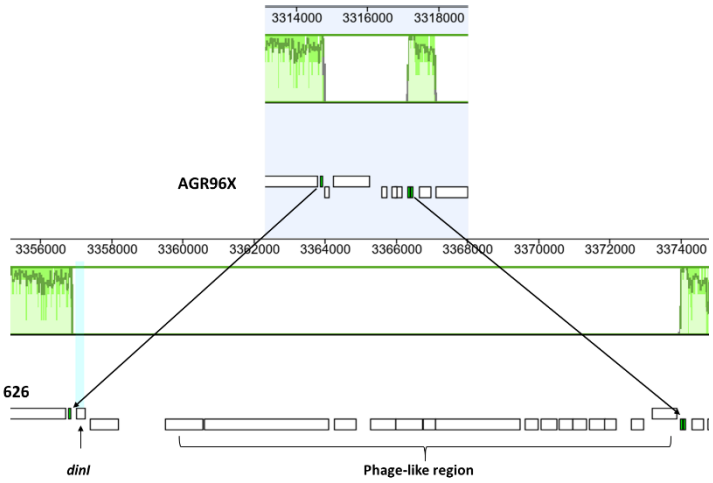


Figure 39 Flanked phage-like region found in *Serratia entomophila* 626
Isolate homogeny of tRNA flanked region (denoted with arrow) between AGR96X (top) and 626 (bottom). Homogenous regions (green) found 5' and 3' of 17 genes phage-like region in *S. entomophila* encoding *dinI* (highlighted in blue).

To help define a possible function of the *dinI,* a spectinomycin antibiotic cassette was used to mutate the *dinI* gene in this phage-like region (Chapter 2, section 2.4.7). Assessing the resultant 626::DinI mutant through bioassay (induction method section 2.3.8 and bioassay protocol section 2.6.2) against *C. giveni* larvae, no pathogenic difference was observed, and no alteration in induction was apparent with MitC when compared with inducible AGR96X (Figure 40). This implies that a *dinI* damage-inducible pathway is not utilised to induce the *S. entomophila* Afp.

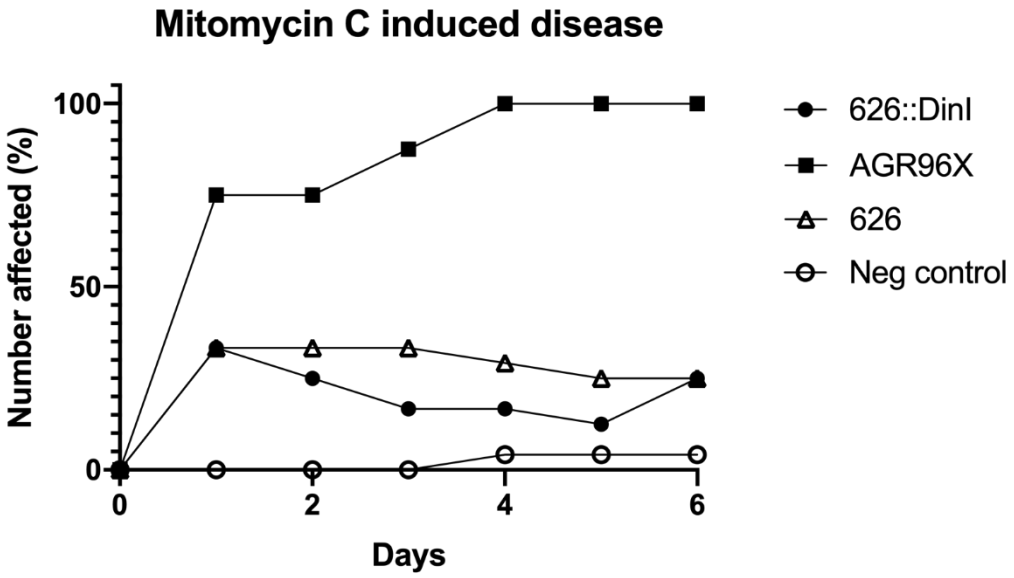


Figure 40 MitC induced expression of virulence factors in *Serratia entomophila* 626 and DinI::spec mutant.
The number of affected grub displaying disease phenotype displayed as a percentage. The negative control showed limited change over the six-day assessment period of induced ultra-centrifuged virulence factors. *S. proteamaculans* AGR96X used as a positive inducible control and *S. entomophila* 626 as the WT equivalent of 626::DinI mutant.

5.2.9 Acquisition of foreign DNA

To determine any probable horizontal gene transfer events on the chromosome, genomic islands were predicted and visualised using Islandviewer4 (2.5.12). It was hypothesised that pathogenic isolates would have more regions of predicted horizontal gene transfer, with non-pathogenic isolates remaining more conserved. For this reason, similar *S. entomophila* isolates 477 (non-pathogenic) and 626 (pathogenic) were used to display comparative genomic islands (Figure 41A and Figure 41B). Genomic island predictions for all isolates not displayed below can be found in Appendix C.8. Comparing both isolates, the pathogenic isolate 626 has more frequent regions of variation on the chromosome when compared to 477. This is complemented by the large spikes in G+C skew seen on

the inner ring at these regions. A salient finding is the presence of the itaconate degradation operon, which is predicted as a genomic island (later outlined in section 5.2.11). Itaconic acid utilisation is used as the plate assay selection method for *S. entomophila* to separate from isolates of *S. proteamaculans*. BLASTn of the five gene genomic island regions from *S. entomophila* isolate 626 identified no significant nucleotide similarity to other species using NCBI BLAST function, in addition to custom BLAST on the cohort of sequences used in this study. This indicates this region is unique to *S. entomophila* with no known origin species. Full annotations of the predicted genomic islands found in isolate 626 can be seen in Table 15.

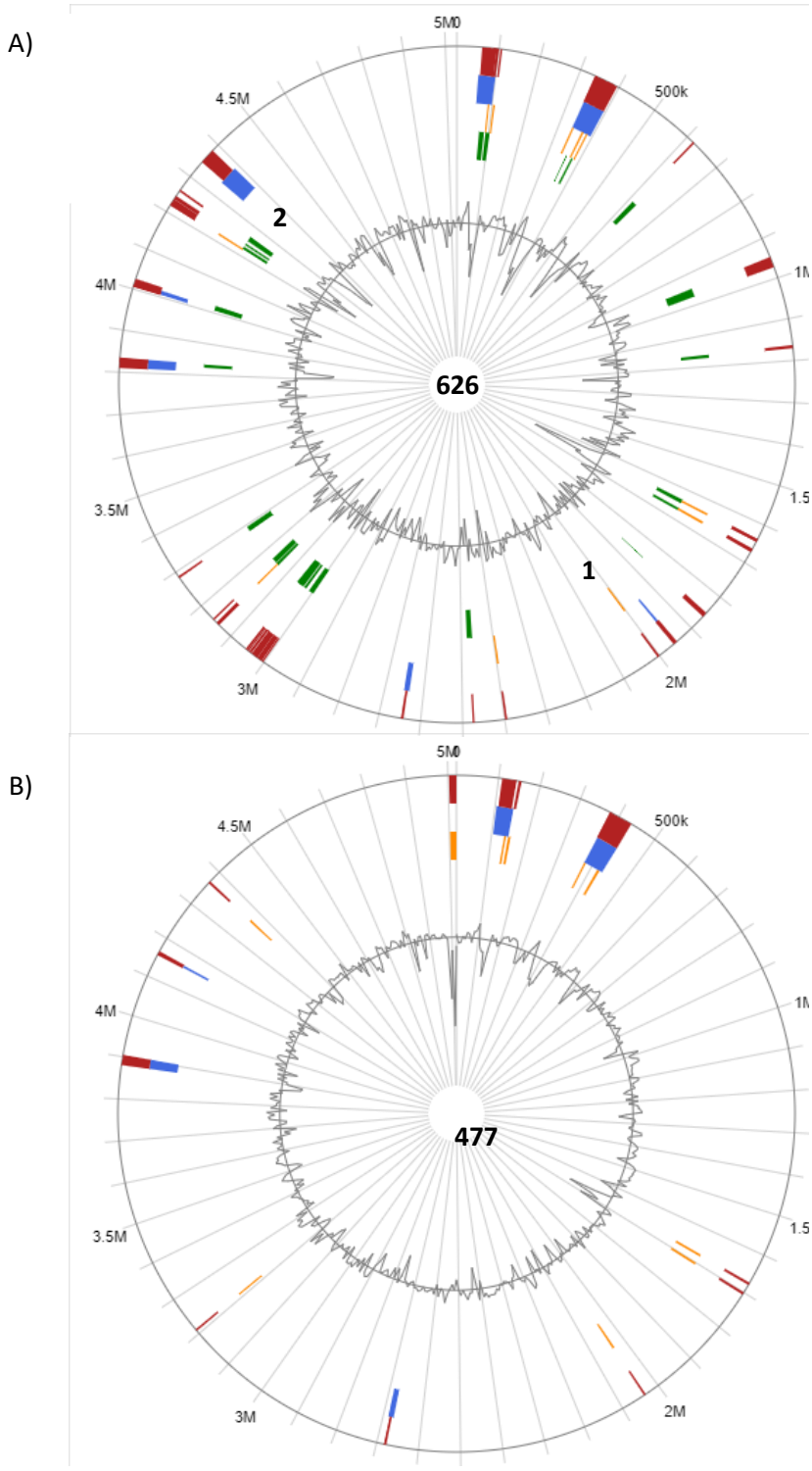


Figure 41 Predicted genomic islands in *Serratia entomophila*
Panel A) shows isolate 626 and panel B) shows isolate 477. Red indicates where a genomic island has been predicted by one of the identification tools utilised by IslandViewer (IslandPath-DIMOB, SIGI-HMM, IslandPick, Islander) where blue, orange and green, respectively, represent alternate prediction tool.
[1] Indicates the itaconate degradation operon.
[2] Indicates the prior described unique 40 Kb region (section 6.2.5)

Table 15 IslandViewer4 hits predicted in *Serratia entomophila* isolate 626.
Predicted genomic islands in isolate 626 by any one prediction method corresponding with Figure 41. Prediction methods are Islandpath-DMOB, IslandPick, SIGI-HMM.

Start-end	Size	Genes of interest
59,894- 94,286	34,392	Prophage integrases and fimbrial units
337,872- 395,069	57,197	Extracellular serine protease *bvGS* virulence sensor Located after enterobactin cluster
617,886- 622,710	4,824	Fimbrial proteins
947,141- 973,038	25,897	Heme binding protein, hypotheticals
1,163,105-1,170,759	7,654	Type II secretion system
1,636,735- 1,643,253	6,518	D-inositol 3-phosphate glycosyltransferase
1,663,337- 1669,591	6,254	O-antigen transporter
1,849,825- 1,861,739	11,914	Outer membrane proteins, transcriptional activators
1,946,689- 1956,596	9,907	Ribonuclease, DinI, phage proteins
2,000,067- 2,005,235	5,169	Itaconate operon
2,393,913-2,398,846	4,933	*rhsD/rhsC*, hypotheticals
2,472,022- 2,476,346	4,324	Hypothetical
2,641,221- 2646,854	5,169	PSP phage shock operon
3,002,010- 3,006,436	4,426	*pipB2* effector protein, *vgrG1* actin cross-link toxin, motility protein, hypotheticals, PrpC
3,006,010- 3,019,240	12,628	*vgrG1, clpV1*, hypotheticals
3,019,257- 3,035,583	16,326	*umuC/umuD*
3,139,939- 3,144,128	4,189	Transcriptional regulators and hypotheticals
3,154,528- 3,158,829	4,301	*holE*_2 DNA polymerase III, hypotheticals
3,283,431- 3,288,130	4,709	HTH-type transcriptional activator, repressor and regulator, ribose import binding proteins *rbsAB*C
3,810,072- 3,834,662	24,590	Partial tox-antitox *pasI*, phage unit. ribose toxin *ratA*
4,007,023- 4,014,342	7,319	Fimbrial genes and *xerD* copies
4,217,383- 4,234,280	16,897	*dinB*, *pksN*, hypotheticals
4,256,747- 4,261,506	4,759	Type I secretion system
4,334,264- 4,281,552	37,288	Phage unit., x*erD*,

To investigate whether any acquisition differences are noted on the geographically isolated Chatham Islands, both 440 and 442 were assessed and compared to *S. entomophila* 626 for differences of predicted genomic islands. Some similarities were observed; the 57,197 bp island was still observed after the enterobactin cluster in 440 but lacked the extracellular serine protease found in 626. The *pspABCD* phage shock operon was also found on 440. Otherwise, the 440 genomic islands encoded unique colicin units and noticeably high copy numbers of transposases from Is3, Is4, and IS630 family transposases (shown in Table 16).

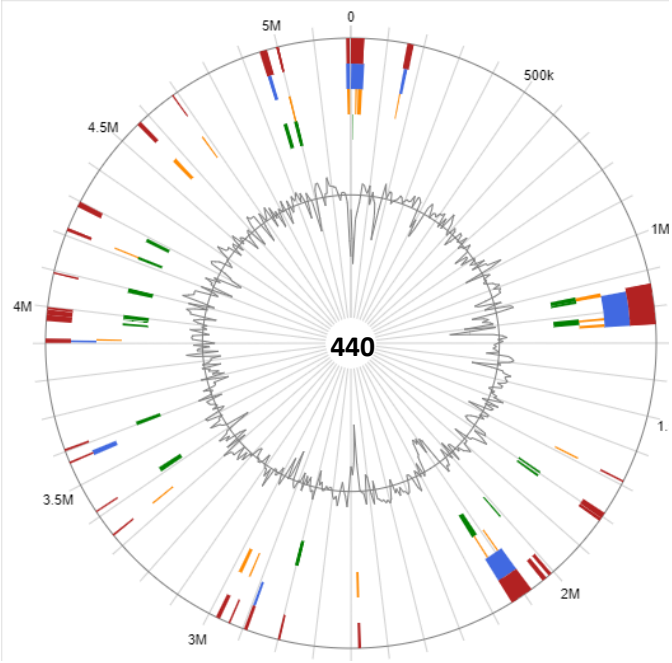


Figure 42 Predicted genomic islands in *Serratia entomophila* isolate 440
Red indicates that the island was predicted by at least one method, where blue (Islandpath-DMOB) green (IslandPick) and orange (SIGI-HMM) indicated individual prediction methods. G+C skew shown in the centre circle.

Another factor to consider would be the lateral transfer of transposases as evidence of horizontal gene acquisition. Taxa that share a transposase are likely to be connected by an ecological niche, transmitting horizontally or vertically. As copies of IS4 ISCro6 were found in isolate iDIA during Roary analysis and confirmed using ISfinder, genome sifting for similar transposases began with this group. Other chromosomes that showed the presence of IS4 IsCro6 transposases in one or multiple copies are in isolates 482, 477, 398, A1M02, 440, 305, MC2, Sent1, and Man3. Isolate 305 and Man3 show IS4 Cro6 in conserved regions implying a common ancestor. The variability of 398, 477, 440, iDIA, and 482 regions around IS4 Cro6 allude to horizontal gene transfer. Comparisons of chromosomal locations of IS4 Cro6 between *S. entomophila* isolates show large levels of similarity between 440 and

MC2, showing ten corresponding locations with flanking genes, implicating vertical transfer of DNA of historical acquisition for these two isolates. This would imply that the relationship of these two isolates, though geographically distant, is a great deal closer than expected. A common flanking gene within the *S. entomophila* chromosomes is *pdeD*, located 3' of IS4 Cro6 in 440, 442, MC2, and iDIA. This gene is a cyclic phosphodiesterase with 86.1% similarity to the *S. ficaria* gene *ycgG* (accession A0A240C3U7). Both 477 and 398 share a similar IS4 Cro6 genome placement, 3' of a gene that showed 100% identity to L-serine dehydratase from *Helicobacter pylori*, which is implicated in carbohydrate biosynthesis (accession Q9ZMU7). Other transposases found on the chromosomes of *S. entomophila* isolates from North and South Island of New Zealand are listed in Table 16, alongside common transposases of Chatham Island isolates. Most show conserved flanking regions, indicating acquisition from a common ancestor. Eleven extra transposases were found on 440; one unique and 10 common with 442, also from the Chatham Islands. Isolates, where HGT events may have occurred, are mostly non-pathogenic (440, 442, 482, 477) with only iDIA as an identifiable candidate for HGT acquired transposases for a virulent isolate. West Coast isolate MC2 shares swathes of commonalities with the three Chatham Island isolates 440, 442, and 294 (Table 16). French isolate 220 does not seem to differ from New Zealand isolates in what transposases are encoded on the chromosome. Encoding a similar transposase profile to other *S. entomophila* isolates implies that this isolate is similar in origin. The diversity shown between the Chatham Island isolates and those of the North and South Islands (which are geographically closer) supports this. When comparing the transposase encoding profile IS3 ISI1661 and IS3 ISRso11 are encoded exclusively on South Island isolates, while the North Island (Waikato) isolate 305 encodes two unique transposases not found on any other isolate. Of the two Southland isolates, only isolate 1343 encodes transposases unique to itself, whereas 364 has a profile in common to other South Island isolates.

Table 16 Presence of predicted transposases found on the chromosome of *Serratia entomophila* isolates using ISfinder displayed as a visual alignment

* Indicates pathogenic isolate. Black indicates the presence of specific transposase.

5.2.10 Accessory virulence factors in *S. entomophila*

To identify potential accessory virulence factors on the chromosomes of *S. entomophila* isolates, HMMER.3 was utilised in parallel with Pfam alignments of family protein domains to construct HMM profiles (Chapter 2, section 2.5.13). Using protein sequences of individual isolates, these profiles were searched to predict the presence of associated proteins, which were then validated by protein searches on UniProt (UniProt 2008).

For searches to locate chitinase enzymes, family *glyco-hydro-18,* and *glyco-hydro-19* domain motifs were predicted using hmmsearch. Chitinases are split into these two families based on amino acid structure, and although most bacterial chitinases are identified from family 18 glycoside hydrolases, both were searched.

5.2.10.1 *In silico* chitinase detection

First, chitinase proteins were defined on the PacBio sequence for isolate 626. Five domains were found, which were then BLASTp 'ed in UniProt to confirm. From this analysis, Chitinase A, Chitinase B, and Chitinase A1 proteins were present in all *S. entomophila* isolates assessed. Of the other two glycoside hydrolase family 18 domains identified on isolate 626, both were listed as hypotheticals. Post BLASTp (Table 17) one was then assigned the protein ChiD with close orthologue in *S. ficaria*, while the other has a lower affinity with sequences in the protein database and was defined as a putative lipoprotein (Table 17). The amino acid sequence of this protein was inputted into Phyre.2 software, where the conformation of the molecule was confirmed with 100% confidence of hydrolase functionality. This approach was conducted for all *S. entomophila* associated searches and generated no differences in encoded predicted chitinase motifs (full outputs shown in Appendix C.9). The same search was conducted for *glyco-hydro-19* domain motifs and returned no results. This was expected, given unlike family 18 chitinases which are broadly found across bacteria, family 19 chitinases are typically found in plants and limited in bacteria (Kawase et al. 2004). Putative signal peptides located on *chiA* and *chiB* were then predicted using SignalP 5.0. A signal peptide cleavage site was identified on ChiA protein between AA positions 23 and 24 with a probability of 0.88, implicating the protein product as a secretory protein via a type 1 secretion system. Uniform prediction of *S. entomophila* encoding *chiA* could define the homogeneity seen here prior in the chitin degradation plate assay responses of *S. entomophila* wildtype isolates Chapter 4, section 2.3.2.6.

Table 17 HMMER results for chitinase glyco-hydro family motif search and chitin-binding protein family on *S. entomophila* isolate 626, with percentage similarity of protein sequence to the most similar protein amino acid sequence on UniProt.

Protein name	Function	Size (AA)	% similarity	Accession
ChiA	Chitinase A	563	97.2% *S. ficaria*	A0A240A944
ChiB	Chitinase B	499	94.4% *S. ficaria*	A0A240C7E2
ChiA1	Chitinase A1	426	86.2% *S. nematodiphila*	A0A086GDZ3
ChiD	Chitinase D	481	91.8% *S. ficaria*	A0A240BK97
'Hypothetical'	Putative lipoprotein	423	67.8% *Yersinia pestis*	Q0WDY9
GlcNAc Binding protein A			97.0% *S. ficaria*	A0A240C8H0

By searching the CAZY (carbohydrate-active enzymes) database, reference genome *S proteamaculans* 568 could be used to find genes of interest as no reference for *S. entomophila* sequences exists currently in the database. It was found that 568 encoded for several chitin-binding proteins on the genome (family AA10/cbm33). Pfam family *LPMO_10* was identified by using the BLAST function on Uniprot to determine Pfam reference. This family was used to build an HMM search profile for chitin-binding proteins. As shown in Table 17, one chitin-binding protein was identified on the genome of 626 and subsequently all other *S. entomophila* isolates. This was annotated by Prokka as *gbpA*, although re-identified through Uniprot as GlcNAc binding protein.

5.2.10.2 Detection of encoded lipases

The same process was utilised for encoded lipase families. Of the five lipase family alignment groups found on Pfam, those for family *Lipase* and family *Lipase_3* were used in this study as the broadest family groups that would cover lipases found in *Serratia* spp. Results from these two family groups were then compiled and duplicate motif results were omitted to formulate an accurate list of proteins identified (full outputs shown in Appendix C.9). The majority of lipases identified from HMM motif searches in isolate 626 had high similarity to that of *S. ficaria* described lipases, demonstrating these lipases are found in other members of the *Serratia* genus (Table 18).

An extracellular phospholipase was found uniform across all pathotypes of *S. entomophila* with high amino acid similarity to an extracellular lipase in *S. liquefaciens.* Bacterial mechanisms for the expression of this phospholipase are mediated by the *flhD* master regulator (Givskov and Molin 1993). Previous studies in *E. coli* show the absence of the polyketide pathway including PhlA and PhlB allows the build-up of PhlA within the cell. This forms a complex with PhlB which inhibits phospholipase

activity to form an inactive complex. All isolates of *S. entomophila* in this study included this polyketide biosynthetic pathway for secretion, corroborating the findings in Chapter 4, which showed no difference in halo size between isolates in lipase assays.

Table 18 HMMER results for lipase family motif searches on *Serratia entomophila* isolate 626, with percentage similarity to most similar protein on UniProt

Protein name	Function	Size (AA)	% similarity	Accession
Monoacylglycerol lipase	Hydrolase activity	268	87.7% *Serratia ficaria*	A0A240AUF3
2-succinyl-6-hydroxy-2,4-cyclohexadiene-1-car	Hydrolase activity	335	93.4% *S. ficaria*	A0A240C062
Hypothetical		614	87.8% *S. ficaria*	A0A240BYY2
Extracellular phospholipase A1		320	90% *S. ficaria*	A0A240CAJ3
Esterase YbfF		257	96.5% *S. ficaria*	A0A240BKF7
Enterobactin synthase component F		1314	96.3% *S. ficaria*	A0A240AIH7
2-succinyl-6-hydroxy-2,4-cyclohexadiene-1-car		256	89.1% *S. ficaria*	A0A240C963
2-(acetamidomethylene)succinate hydrolase		284	75.4% *S. ficaria*	A0A240CAQ6
High-affinity branched-chain amino acid trans	Part of the ABC transporter complex	316	96.5% *S. ficaria*	A0A240C2G3
Hypothetical	Phospholipase	500	69.9% *Pluralibacter gergoviae*	A0A0J5NHR8
Hypothetical	Alpha/beta hydrolase fold	271	86.0% *S. ficaria*	A0A240C523
S-formylglutathione hydrolase YeiG	Serine hydrolase	280	98.2% *S. ficaria*	A0A240BR47
Carbonic anhydrase 1	Hydration of CO_2	217	80.7% *Achromobacter piechaudii*	D4X5X4
Putative aminoacrylate hydrolase		261	82.6% *S. ficaria*	A0A240BVH3
Carboxylesterase NlhH	Lipase 2 triglyceride	324	95.1% *S. ficaria*	A0A240C634
Non-heme chloroperoxidase	Peroxidase	276	73.2% *Microvirga vignae*	A0A0H1R2S0
Hypothetical	Predicted esterase	231	90.4% *S. ficaria*	A0A240C5H0
Aclacinomycin methylesterase RdmC	Enol-actonase activity	307	86.4% *S. ficaria*	A0A240C3Z7
Acetyl esterase	Lipase 2	291	94.5% *S. ficaria*	A0A240C087
Putative aminoacrylate hydrolase		272	90.8% *S. ficaria*	A0A240CAA2

5.2.10.3 Detection of encoded extracellular DNases

As outlined in Chapter 4, the higher levels of DNase secretion in *S. entomophila* isolates could potentially limit the acquisition of horizontally acquired DNA, leading to a general reduction in variation. A literature search was conducted to find relevant genes that could be associated with the extracellular secretion of nucleases in *Serratia* spp. For example, *nucA* is a well characterised gene of *S. marcescens* (Suh et al. 1996). As this is a well-defined secreted nuclease, BLASTp function of UniProt was initially used to identify *nucA* or similar genes in 626 from HMM pFAM searches (PF14040). A second endonuclease, NucM, was identified that is closely related to EndA in *S. ficaria*. This protein shares 75.3% amino acid identity with *Dickeya dadantii* NucM. Uniprot described this gene as potentially implicated in protecting the bacteria against foreign DNA, with the potential to be involved in pathogenicity via a signal sequence-dependent pathway.

On investigation of the complete dataset of 23 *S. entomophila* chromosomes, all isolates encoded a single *nucA* and *nucM* copy with homology to those for *S. marcescens* and *S. ficaria* in GenBank (Table 19). Plate assays conducted in Chapter 4 show that *S. entomophila* isolates of varying pathotypes have a similar range of DNase expression, which corroborates the finding of similarly encoded extracellular nucleases.

Table 19 Identification of extracellular secreted nucleases
Serratia entomophila isolate 626, and its closest % similarity from UniProt BlastP function

Protein name	Function	Size (AA)	% Similarity	Accession
NucA	Catalyses the hydrolysis of both DNA and RNA	266	94.4% *S. marcescens*	P13717
NucM	Endonuclease	232	95.2% *S. ficaria*	A0A240CAN3

5.2.10.4 Presence of peptidases and extracellular proteases

The peptidase database MEROPs was used alongside the NCBI BLAST database to create a library of peptidase motifs to search the *S. entomophila* chromosomes for encoded proteases. Over the 23 isolates of *S. entomophila* assessed, a mostly uniform cohort of peptidases were predicted. From the literature and MEROPs database, *S. proteamaculans* 568 encodes 123 peptidases, wherein 43 are homologous proteins. In contrast, *S. marcescens* encodes for 306 peptidases of which 111 are homologs.

S. entomophila isolate 626 is predicted to encode 132 peptidase variants (full results in Appendix file D.6). Overall, the range of predictions made by MEROPs is from 126 (isolates iDIA, 210, 158) to 163 (isolate 364) variants. Most of the peptidases and their homologues predicted are found across multiple *S. entomophila* isolates, although three peptidases are unique to isolate 440, two unique to iDIA, and one each to isolates 626, MC2, and 1100. Seventy-three predicted peptidases are found across all *S. entomophila* isolates. The full presence/ absence file created for the MEROPs peptidase hits can be found in Appendix D.6. Of the differences observed and corresponding with BLASTp of genes found using MEROPs searches, 626 encodes for five extracellular serine proteases on the chromosome. Both Chatham Island isolates (440 and 442) only encode for a single serine protease. Only 626, 345 (encodes six) 158, and 176 have more than three. The chronic pathogen iDIA encodes for three of these extracellular serine proteases. Additional extracellular proteases are found exclusively on pathogenic isolates, though are not inclusive of the whole pathogenic cohort of *S. entomophila*. This result shows that there is potential for extracellular proteases to contribute to the pathotype of disease in *C. giveni* larvae. Extracellular and metalloproteases identified in isolates of *S. entomophila* have all been previously described in other *Serratia* spp. (Table 20).

Table 20 Extracellular serine proteases and metalloproteases predicted in *Serratia entomophila* isolate 626
AA sequences BLASTp against Uniprot database for closest related protein homolog.

Protein name	Function	Size (AA)	% Similarity	Accession
Autotransporter protein	Outer membrane transport	1,145	62.5% *Izhakiella* sp.	A0A4P8YMK4
Extracellular serine protease		1,145	71.8% *S. marcescens*	A0A240C7V3
Extracellular serine protease		1,031	90.2% *S. ficaria*	A0A240CAN3
Extracellular serine protease		1,021	94.4% *S. ficaria*	A0A240CDN3
Extracellular serine protease		1,004	94.0% *S. ficaria*	A0A240CDP9
ATP dependent zinc metalloprotease FtSH	For cytoplasmic and membrane proteins. Transport of toxic c-terminal of *Y. pestis* in infection Putative metalloprotease	646	99.5% *S. ficaria*	A0A240AS00
Metalloprotease *yggG* (loiP)	Metalloprotease		96.0% *S. ficaria*	A0A240CA27
Peptidase PmbA	Metalloprotease		98.2% *S. ficaria*	A0A240CBL7
Protease TldD			96.9% *S. ficaria*	A0A240CBP3

5.2.10.5 Prediction of biosynthetic clusters

Chromosomes were screened for the presence of accessory metabolic and virulence factors to determine any distinct differences that could enhance bio-infectivity or aid an isolate of *S. entomophila*. Using antiSMASH across 23 chromosomes the predicted secondary metabolites are generally conserved with small differences between groups of isolates. Of the 6-10 predictions per isolate, six metabolites were uniform across all isolates, including two non-ribosomal peptides (NRPS), a beta lactone, a siderophore, and a thiopeptide (Table 21).

Of the unique metabolites predicted, isolates MC2 and 440 have a ladderane, a phospholipid compound (Table 21). BLASTp of the core biosynthetic gene (*fabF*) region reveals this had 99.0% similarity to beta-ketoacyl synthase II family protein from *S. ficaria*. AntiSMASH prediction of this cluster shows the prediction of *fabF* as the core biosynthetic gene, with *plsC_2* (phosphate acetyltransferase), *gdhI_3,* and *itnD* as accessory biosynthetic genes. Both *fabF* and *plsC* have previously been implicated as ladderane biosynthetic genes in *E. coli* (Javidpour et al. 2016). These two *S. entomophila* isolates (MC2 and 440), although geographically distant, have a close phylogenetically inferred relationship. However, it is important to note that only MC2 and 440, and no other Chatham Island, isolates share this genomic trait.

When comparing to similar known clusters using antiSMASH KnownClusterBlast function, a cut-off point of 60% was used to determine proximity to a known cluster to establish a potentially functional relationship. Uniform across all isolates of *S. entomophila*, siderophore predictions shared 77% similarity. An aerobactin siderophore has been previously described in *Xenorhabdus*, implicated in the virulence of invasive pathogens. For example, aerobactin has shown pathogenic properties to *Galleria mellonella* (Lepidoptera), with mutants expressing delayed infection, implicating the cluster in initial colonisation and establishment (Hirschmann et al. 2017). Although this cluster was found uniformly across all pathogenic and non-pathogenic isolates, there is the potential that aerobactin performs similarly in aiding the establishment of the pathogen post-challenge. A terpene cluster found in 10 isolates showed high similarity (75%) to a sodorifen synthesis cluster described from *S. plymuthica* (Domik et al. 2016). Sodorifen production is promoted in response to fungal volatile organic compounds in *S. plymuthica*, suggesting a system that may aid communication between soil bacteria and fungi and potentially allow the bacteria to respond to changes in the environment by detecting fungal volatiles. The presence of this terpene cluster was not uniform to any pathotype and was present in all of the Chatham Island *S. entomophila*, in addition to West Coast isolate MC2.

Table 21 AntiSMASH analysis for predicted secondary metabolite cluster in all isolates of *Serratia entomophila*.
AntiSMASH results compiled into a presence or absence table for each predicted secondary metabolite cluster. Annotations are as shown fromantiSMASH output.

	Secondary metabolite cluster												
Isolate	Hserlactone	NRPS	Betalactone	Siderophore	Thiopeptide	NRPS	T1PKS/ NRPS	Ladderane	Terpene	NRPS	NRPS	T1PKS	T1PKS
158[1]	+	+	+	+	+	+	+	-	+	-	-	-	-
176[1]	+	+	+	+	+	+	+	-	+	-	-	-	-
210[1]	+	+	+	+	+	+	+	-	+	-	-	-	-
219	+	+	+	+	+	+	+	-	-	-	-	-	-
220	+	+	+	+	+	+	-	-	-	-	-	-	-
305[1]	+	+	+	+	+	+	-	-	-	+	+	-	-
345[1]	+	-	+	+	+	+	+	-	+	+	-	-	-
364	+	+	+	+	+	+	-	-	-	+	-	-	-
398[1]	+	+	+	+	+	+	-	-	-	+	-	-	-
440[2]	+	+	+	+	+	+	-	+	+	-	-	-	-
442[2]	+	+	+	+	+	+	-	-	+	-	-	-	-
477	+	+	+	+	+	+	+	-	-	-	-	-	-
482	+	+	+	+	+	+	-	-	-	+	+	-	-
626[1]	+	+	+	+	+	+	+	-	-	-	-	-	-
1100[1]	+	+	+	+	+	+	-	-	+	-	-	-	-
1343	+	+	+	+	+	+	-	-	-	+	+	-	-
A1M02[1]	+	+	+	+	+	+	-	-	-	+	+	-	-
I[1]	+	+	+	+	+	+	+	-	+	-	-	+	+
iDIA[1]	+	+	+	+	+	+	+	-	-	-	-	+	-
MC2	+	+	+	+	+	+	-	+	+	-	-	-	-
294	+	+	+	+	+	+	-	-	+	-	-	-	-
Man3	+	+	+	+	+	+	-	-	-	-	-	-	-
Sent1	+	+	+	+	+	+	-	-	-	-	-	-	-

[1] Denotes pathogenic isolates
[2] Denotes plasmid bearing non-pathogenic isolates

5.2.11 Itaconate mutant

As the only species of *Serratia* able to utilise itaconate, *S. entomophila* has long been selected on agar media using itaconate differentiate from *S. proteamaculans* (O'Callaghan 1993). The region encoding itaconate utilisation was previously implicated in section 5.2.9 as a predicted genomic island. As described by Hersch and Navarre (2019), the itaconate degradation operon is comprised of four genes, three of which (*mcl, mcH, yfdE*) have homologs previously implicated in itaconate degradation. Previous literature has shown two recorded operon organisations, one with six genes from *Pseudomonas aeruginosa* (*ict ich ccl*) and three non-related genes (MmgE-PrpD family protein, acyl-CoA dehydrogenase, and a glyoxalase family protein). The second, the *Y. pestis* operon, comprises three genes as functional homologs to the *Pseudomonas aeruginosa* genes, *ripABC*. Both operons have been implicated in the degradation of itaconate, with *Y. pestis ripABC*-encoding operon implicated in the survival of bacteria in host macrophages (Sasikaran et al. 2014). The identified itaconate utilisation operon in *S. entomophila* is most similar to the *Y. pestis* three gene operational system of *ripABC*, identified as *mcl, mch and yfdE* when annotated in this study (Figure 43). This operon is 3' adjacent to a *hcaR* regulator gene on the reverse strand. *HcaR* has been previously implicated as a translational inducer of expression. The final 3' gene of the operon functionally is unknown, although Phyre.2 results suggest that this is a transportation protein of the NADC family with 90% similarity to the protein sequence. This region is uniform in its composition across all 23 isolates for *S. entomophila*; however, in isolates 220, 305, 1343, and 364 the operon is in the opposing direction. This suggests acquisition of the operon via horizontal gene transfer. *Serratia ficaria* isolate 457 (which is placed with *S. entomophila* isolates in 16S analysis but differed in ANI) was assessed for the presence of this cluster which revealed no gene involved in this pathway was present. The operon in *S. entomophila* is flanked 5' by a *pip* gene and 3' by *holE*. In *S. ficaria* 457, these two flanking genes were represented, supporting the conclusion that the itaconate degradation cluster is a horizontally acquired region.

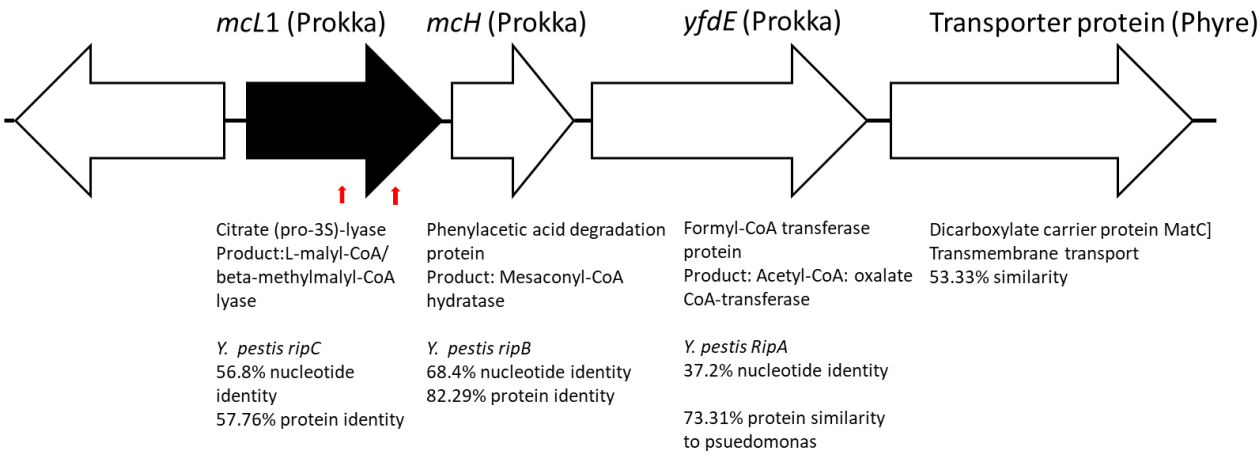


Figure 43 Schematic of the *Serratia entomophila* itaconate degradation pathway operon.
(Left to right) Outside of the operon regulator protein. Mcl, Mch, YfdE, and a hypothetical protein. Red arrows denote the presence of the two NcoI cut sites that allowed for the insertion of the tetracycline cassette, which removed 250 bp and inserted 1500 bp into the sequence, interrupting the Mcl gene, to form the Mch1::tet mutant.

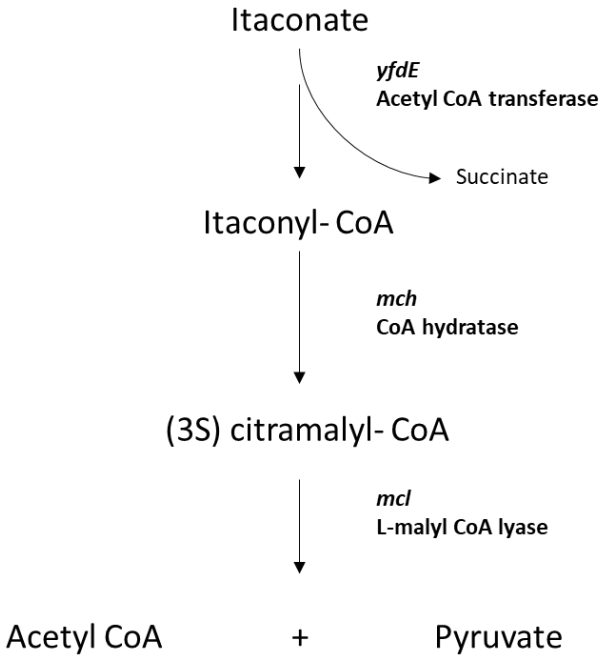


Figure 44 Predicted degradation pathway utilised in *Serratia entomophila*, modified from Sasikaran et al. (2014) and MetaCyc pathway for itaconate degradation.
Genes predicted in *S. entomophila* are shown highlighted and italicised with their relative function. Predicted compounds and products derived from literature and Prokka/BLASTp prediction of relative gene functions.

Based on the current literature and the predicted functionality of genes involved in the itaconate degradation cluster in *S. entomophila* (Table 22), a reconstruction of the predicted itaconate degradation pathway based on the functional homologs of the predicted genes was developed (Figure 44). The first enzyme of the pathway, a CoA transferase encoded by *yfdE*, can be utilised by multiple CoA donors. Previous studies describe various transferases that can be utilised as homologs in this step. *Mch*, the second pathway encoded enzyme, catalysed the conversion of itaconyl CoA into (3S) citramalyl- CoA. As previously reported, only closely related enzymes of the (3S) malyl-CoA enzymes can be utilised by this stage of the reaction. The (3S)-citramalyl-CoA is then cleaved by *mcl* resulting in Acetyl CoA and pyruvate. Synthesised pyruvate can then be utilised by bacteria in the Krebs (citric acid) cycle by generating Acetyl-CoA by hydrolysis to form an energy supply.

Table 22 Closest BLASTX hits to genes located in the *Serratia entomophila* itaconate degradation operon and associated transcriptional regulator

Protein	Size (bp)	Putative Function	BLASTX hit	Accession
HcaR	897	Transcriptional regulator	MULTISPECIES: LysR family transcriptional regulator [unclassified *Pseudoxanthomonas*] 96% query cover 66.21% identity	WP_130522517.1
Mcl	834	CoA ester lyase	[*Bradyrhizobium erythrophlei*] 91% query cover 63.53% identity	WP_079567479.1
Mch	528	Mesaconyl-CoA hydratase	MULTISPECIES: MaoC family dehydratase [*Burkholderia*] 99% query cover 84% identity	WP_099317990.1
Yfde	1,179	CoA transferase	MULTISPECIES: CoA transferase [unclassified *Pseudomonas*] 97% query cover 74.81% identity	WP_008146597.1
Transporter protein*I	1,293		MULTISPECIES: hypothetical protein [unclassified *Pasteurellaceae*] 94% query cover 56.62% identity	WP_139612570.1

* Indicates where function ascribed by Phyre protein sequence search

In five of the *S. entomophila* isolates relative to the chromosome this operon is reversed (Figure 45). No universal phenotype corresponds to the isolates reversed. Isolate 220 was isolated in France and the others in New Zealand. Pathogenicity also varied, with 305 and 364 being pathogenic towards grass grub larvae. They are, however, co-located on the same subclade of *S. entomophila* based on

MLSA performed in section 5.2.3. The reversal of this operon and its regulator in some isolates is another clear indication that this region has been acquired via horizontal gene transfer.

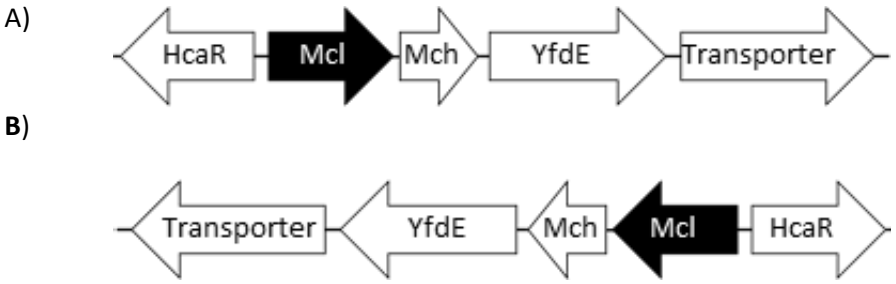


Figure 45 Chromosomal orientations of the itaconate degradation operon
A) 158, 176, 210, 219, 345, 398, 440, 442, 477, 626, A1M02, I, iDIA, MC2, Man3
B) 220, 305, 364,482, 1343

To test for similarity and potential evolutionary history, phylogenetic assessment of the *Y. pestis ich* gene and *P. aeruginosa* orthologue constructed against the predicted *mch* gene in *S. entomophila* was undertaken. Amino acid alignments showed high degrees of differentiation between *S. entomophila* and *P. aeruginosa*, but the protein sequence for *Y. pestis* showed higher levels of sequence homology.

Figure 46 Phylogeny of Mch amnio acid sequence of the itaconate degradation cluster *in Serratia entomophila* with protein homologs
Protein accession numbers for reference sequences are found in brackets. Tree rooted to *Pseudomonas aeruginosa* Ich homolog. Performed with 100 bootstraps using a neighbour-joining tree in MEGA X.

5.2.11.1 Assessment of an itaconate mutant

An itaconate mutant was constructed by inserting a tetracycline antibiotic cassette into the double *Nco*I enzyme cut point (depicted in Figure 43) following methods described in Chapter 2, section 2.4.8. The mutant was sequence-validated after identification by patching onto itaconate agar. A *trans* complement of the *mcl* gene was then tagged and reinserted into the Mch1::tet mutant (section 2.4.8) and sequence validated, before being tested against Mch1::tet and wildtype 626 when patched onto LB+Cm and itaconate plates (Figure 47).

As demonstrated in the literature, there is evidence that the itaconate degradation cluster could be utilised as a metabolic cluster beneficial to *S. entomophila* in its niche. To elucidate any correlation, a 48 h growth curve for both WT *S. entomophila* 626 and a constructed Mch1::tet mutant (section 2.4.8) were conducted in LB and M9 minimal salts (glucose) to determine any metabolic benefit.

Under ideal conditions in LB broth, little difference can be observed in the growth of WT 626 and the itaconate mutant (Figure 47). Trends observed in M9 minimal medium, however, show the growth of Mcl1::tet is potentially impacted by the lack of functional degradation cluster immediately in the exponential growth phase, though growth evens out by 48 h time points. Overall, not much difference can be seen between the two isolates in relation to colony growth levels. It was predicted that the itaconate degradation pathway in *S. entomophila* would allow for the synthesis of pyruvate by a less energy-consuming pathway. Other pathways to generate pyruvate however remain encoded on the chromosome, therefore, any differences between the mutant and wildtype may be negate by compensation utilising other metabolic modules.

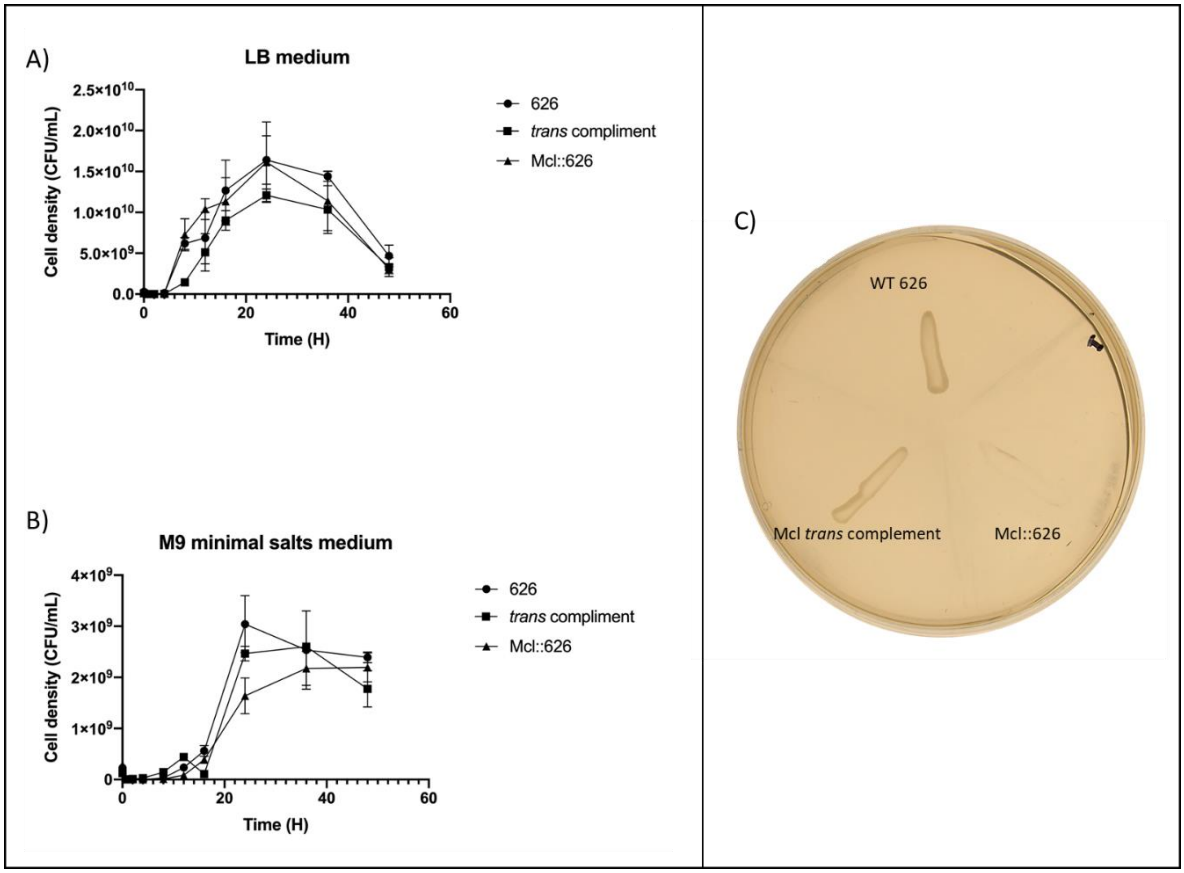


Figure 47 Growth curves of WT 626, Mcl1::tet and complemented *mcl1* gene in optimal and stress conditions

48 h growth curves in triplicate with standard error shown. A) shows growth in LB media B) in M9 minimal media C) shows growth of wildtype 626 against the mutt Mcl::626 and *trans* complemented Mcl gene on itaconate agar.

5.2.11.2 Bioassay of *S. entomophila* Mcl1 mutant against grass grub larvae

Lethal concentration (LC_{50}) and lethal time (LT_{50}) are good indicators of any effect a mutant such as Mcl1::tet can have on the infectivity of a bacterium. Through assessment of triplicate bioassays of *C. giveni* larvae challenged with isolates 626 or Mcl1::tet defined a respective LC_{50} of $4.1 \times 10^6 \pm 9.8 \times 10^5$ CFU, to $2.0 \times 10^6 \pm 1.4 \times 10^6$ CFU (Table 23). The difference observed is minimal with the overlapping range of the standard error, although the mean concentration dose for 50% infection with 626 is approximately 10-log fold lower than that observed in the itaconate pathway mutant. Results of these bioassays suggest that the interruption of the itaconate degradation pathway has a nominal, if any effect, on bacterial pathogenesis. This result may be expected as the predicted function of this cluster points to use in bacterial carbohydrate metabolism. It can be hypothesised that metabolic impacts will likely reduce the ability of a bacterium to colonise a host.

Table 23 Virulence for mutant Mcl1:tet and the wildtype control of *Serratia entomophila* isolate 626 to grass grub larvae, including LC_{50} and LT_{50} with standard error calculation P values (Fisher's exact), with statistical significance in relation to the negative control highlighted in bold

Isolate	LC_{50} CFU[1] (± standard error)		LT_{50}[2] (days)	Diseased (%) ± standard error		Mortality ± standard error		Combined ± standard error	
Blank		-	-	4.16	± 4.16	16.66	± 7.77	20.93	± 8.4
626	1.9×10^5	± 1.3×10^5	3	70.83	± 9.47**(<0.001)**	29.16	± 9.47 (0.492)	100.00	± 0 **(<0.001)**
Mcl1:tet	2.0×10^6	± 1.4×10^6	4	75.00	± 9.02**(<0.001)**	25	± 9.02 (0.742)	100.00	± 0 **(<0.001)**

[1] Undertaken via Probit analysis

[2] Statistical Survival analysis

5.2.11.3 Competition assays in grass grub larvae

A cube of carrot was inoculated with 5 µL of overnight culture (3.1 ×10^7 CFU for 626 and 4.8 ×10^7 CFU for Mcl1::tet) of both WT and mutant 626 and fed to healthy grass grub larvae (~24) to determine any competitive advantage afforded to either isolate. Larval macerates at days 3, 6, 9, and 12 were plated onto selective antibiotics for the Mcl1::tet insertion directly, whereas CTA was used to select for *Serratia*. Species were then determined by patching onto DNase and tetracycline selective LB agar plates. Macerates were undertaken in triplicate for each time point following the protocol outlined in section 2.6.4, with results shown in Figure 48.

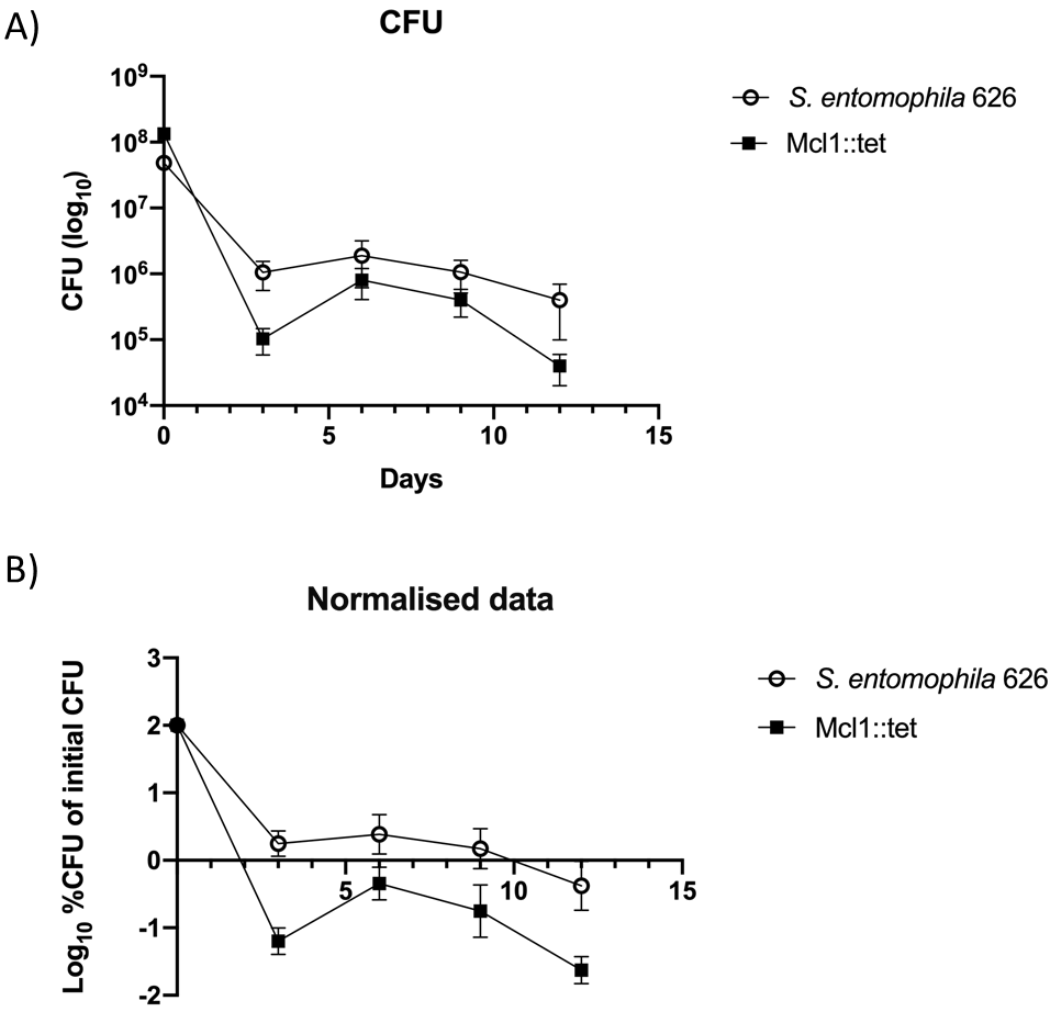


Figure 48 *In vivo* competitive growth experiment.
In vivo growth curves of a 12 day 50:50 competition assay between 50/50 inoculants of WT 626 and itaconate pathway mutant, Mcl1::tet, represented on a log10 scale. A) Raw CFU log_{10} results for each isolate recorded in triplicate for three-day intervals and B) data normalised as log_{10} transformed percentage values against the initial CFU. Error bars represent standard error.

Although initial overnight culture CFU for WT 626 (6.2 ×10^9 /mL) was lower than for the itaconate mutant (9.6 ×10^9 /mL), the CFU re-isolated from *in vivo* macerate samples remained more stable after

the initial drop off from inoculation than Mcl1::tet (Figure 48). Populations for both strains had stabilised by day three (Figure 48). Two sample t-test of the normalised data for day three shows that the mean re-isolation at this time point is statistically significant (P=0.013), where WT 626 shows an advantage in establishment over the mutant. WT 626 had a stable 10^6 CFU load over the 12-day competition assay, which was not observed in the mutant. By day 6, Mcl1::tet population was declining and this decline continued through day 12, both as a percentage in comparison to the wildtype, and overall dropping to mid ranges of 10^4 CFU. While the endpoint difference is not statistically significantly different (P=0.057), WT 626 is trending towards a growth advantage, with a 1 log-fold difference in re-isolation. *In vitro* and *in vivo* analysis observations show the itaconate degradation pathway does confer some metabolic benefit to wildtype *S. entomophila* isolates, where the *mcl* mutant does not achieve the same cell density in the host as wildtype 626 in addition to the slight changes observed in LB and M9 culture in section 5.2.11.1.

5.2.11.4 Protozoa survival assay

It has been noted that *Yersinia pseudotuberculosis* can persist within the trophozoite stage protozoa (Santos-Montañez et al. 2015). In addition to studies noting that the itaconate pathway in other *Yersinia* spp. allows for survival of the bacterium in macrophages, a preliminary investigation was undertaken to determine whether the itaconate pathway in *S. entomophila* had similar beneficial effects for the endemic gut protozoa of *C. giveni*, relative to its ability to survive long term in soil. In this context, it is important to note that the grass grub hindgut specifically the fermentation sack, has a high population of protozoa (Miln 1978).

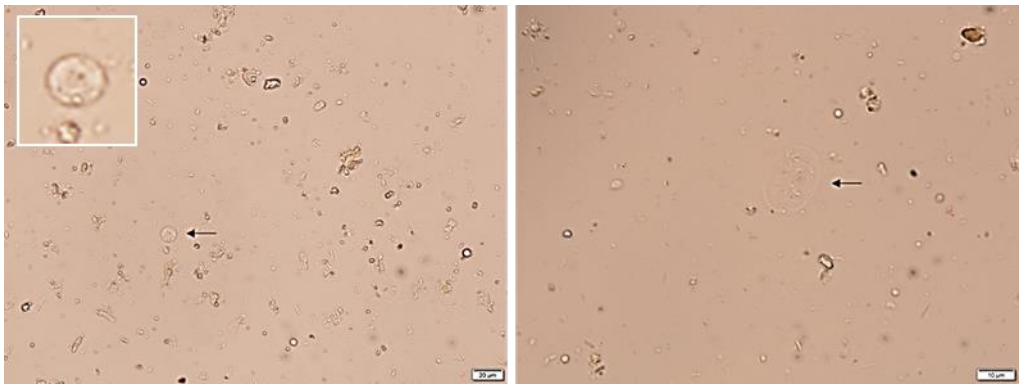


Figure 49 Protozoa (indicated by black arrow refer inset for expanded view) in a 1:10 dilution of larval fermentation sac pooling sample.

To determine any link between *S. entomophila* survival and protozoan symbiosis (Figure 49), samples were taken of *C. giveni* larval gut haemolymph and co-cultured as 1 in 10 dilutions of fermentation sac fluid (purified as outlined in Chapter 2 section 2.6.6) and distilled water with neat *S. entomophila* culture to 200 µL overnight culture (16 h). Samples were incubated anaerobically for 6 h before a

gentamycin wash for 1 h, then lysis of the protozoa with Triton X100. Plating out of the solution was undertaken at three stages (prewash, post-wash, post lyse) on ITA medium (to select for *S. entomophila*). Both gentamycin washed samples and lysed protozoan samples did not have any growth after seven days at all dilutions (10^{-2}-10^{-3}). *Serratia* isolates were expected to grow after 4-5 days of incubation.

Results indicated that although *Serratia* could be isolated from the pre-gentamycin wash step, no *Serratia* was isolated post gentamycin wash or post Triton X-100 lysis induction. This would indicate that *S. entomophila* is not using protozoa as a survival reservoir within the host, as is seen with macrophage survival in *Yersinia* spp. However, this was only a preliminary assessment and additional more precise studies are required, to eliminate the possibility that *S. entomophila* can survive in protozoa. In this context, the ability to purify and obtain sterile, clean, concentrated amounts of protozoan species found in the grass grub fermentation sack would be advantageous and undertaking the experiment under precise anaerobic conditions to replicate the in-host environment more closely.

5.3 Discussion

Through this analysis, a gold standard genome of the *S. entomophila* BioShield® strain 626 has been sequenced and annotated and aspects of the *S. entomophila* chromosomes have been described. Their relation to other *Serratia* species has also been categorised, where a high degree of homogeneity was noted between *S. entomophila* isolates. Phylogenetic analysis of 16S rRNA showed that *S. entomophila* resides in the genus *Serratia*, closely related to *S. ficaria*, an environmental bacteria isolated from the fig tree ecosystem, including fruit and fig wasps (Grimont et al. 1979). Combining the output from 16S in Chapter 3, with MLSA, ANI, and whole-genome comparisons in this chapter show that the separation between *S. entomophila* and *S. ficaria* warrants further investigation to finalise the species position. The dissimilarity between the *S. ficaria* type strain and the in-house isolate *S. ficaria* 457 implies that the delineation of the species boundaries for this isolate needs additional study. It is evident however from MLSA phylogenies that *S. ficaria* is a separate species *to S. entomophila*, but *S. ficaria* 457 may have been misidentified.

As discussed in section 5.2.2, the main difference between the chromosomes of *S. entomophila* isolates is the varied number of encoded rRNA copies. High copies of rRNAs are seen in West Coast isolates MC2 and 219, Chatham Island isolate 440, and Canterbury isolates 626 and iDIA. rRNA operons are generally assembled as units of 16S, 23S, and 5S rRNA, with *E. coli* having seven of these operons copies (Ellwood and Nomura 1980). With the assembly of these operons, all five of the high encoding isolates have seven copies, as seen in *E. coli*. Both Canterbury isolates 626 and iDIA are the only pathogenic *S. entomophila* to encode these additional rRNAs on the chromosome. rRNA copy number has previously been associated with competitive fitness in soil microbes. Reproductive success of high rRNA copy bacterial isolates has been described in competitive microcosm assays by utilising faster response mechanisms to environmental changes (Klappenbach et al. 2000).

Analysis of COG comparisons in relation to *S. entomophila* chromosome 626 shows that several categories are reduced in comparison to other chromosomes of *Serratia* spp., although this can be partially explained by the smaller genome size.

From assessing the chromosomes of *S. entomophila* isolates, several points of evidence suggest that *S. entomophila* has reduced uptake of environmentally acquired DNA compared to *S. proteamaculans* (to be discussed in Chapter 6). Limited numbers of genomic islands were found, though these are usually limited to recent transfer events as historic genomic transfer will, over time, ameliorate to the chromosome. A factor to consider when assessing *S. entomophila* proclivity for horizontally acquired genetic material is the barrier of similarity between the donor and the recipient. A previous study by Popa et al. (2011) recorded that in 86% of prokaryote genomes assessed the donor-recipient G+C

difference was <5%. In the instance of *S. entomophila,* the average genome-based G+C content was 59%. Soil bacteria have been previously described as generally having higher G+C content, which would not explain the lack of acquisition from soil bacteria. However, if *S. entomophila* survives long-term in an insect host, the ability to receive genomic DNA from these soil microbial communities may be limited, but on the other hand, may be increased through proximity to the gut microbiota.

Another factor to consider would be the co-habitation of microbes in an ecosystem. Hooper et al. (2009) provide evidence that lateral transfer of transposases can be used to determine habitat range. In this instance, the detection of transposases and their surrounding gene region can be used to determine whether horizontal gene transfer in an area is mediated. In the case of IS4 Cro6, there are copies implicated as lateral transfer events in 477, 440, 398, and iDIA. The presence of transposases can and has previously been used as a proxy to determine genome plasticity, representing horizontal gene transfer events (Vigil-Stenman et al. 2017). Most isolates of *S. entomophila* encode 0- 2 transposase types. IS5 IS4811 was deduced to be a vertical transfer from a common ancestor due to the conservation of its flanking regions. Taking this into account, the low *in silico* evidence of horizontal gene transfer for transposases is a fair representation of the potential for HGT events integrating novel genomic material into *S. entomophila* isolates. Of the isolates potentially implicated in HGT events, only iDIA was pathogenic.

In plate response assays assessing the production of various accessory virulence determinants (section 4.2.5), the production of accessory factors resulted in no statistically significant change in any category between pathogenic and non-pathogenic isolates of *S. entomophila* except in the case of DNase. . This implicates there are regulatory factors yet undefined that mediate the secretion of these nucleases, potentially controlling the uptake of free DNA in the environment (evidence for this is provided through the assessment of bacterial transconjugants in Chapter 7 and transcriptome analysis in Chapter 8).

These HGT events seem to underline the main regions of differentiation in the chromosomes of *S. entomophila.* Whole chromosomal alignments show that, overall, the chromosomal makeup of the species is uniform. When correlating genome alignments with the prediction of genomic islands most regions of variations align closely to the presence of specific islands on the chromosome. The most variation seen was in Chatham Island isolates, where chromosomal inversions and rearrangements were found. Bacterial genome stability is inferred by the regulation of homologous recombination of inverted repeats. As was described in Chapter 4 very little difference in pathotype, genotype or phenotype was noticed between geographic areas. Contrary to what is observed in other pathogen/ non-pathogen bacterial systems the chromosomes in the Chatham Island isolates (440, 442, and 294)

are larger than their mainland New Zealand counterparts, encoding more coding genes. In the parallel evolution of *Streptococcus pneumoniae* and non-pathogenic *Streptococcus mitis*, the latter displays a stabilised genome of ~15% reduction in size to *S. pneumoniae*. This change is characterised by the absence of many virulence-associated genes (Kilian et al. 2014). It was also found that these isolates encode for different accessory virulence/defence mechanisms compared to the North/South Island counterparts, in addition to small changes in metabolic pathways As was noted in Chapter 4, the West Coast isolate MC2 again shares many common features with the Chatham Island isolates, despite the geographic separation. A ladderane unique to the Chatham Island isolate 440 and MC2 from the West Coast stood out from the usually ubiquitous results from biosynthetic cluster prediction. The nongeographical conformity of these additional biosynthetic groups and the lack of this cluster from Chatham Island isolates 294 and 442 show that this region was probably acquired from an event before the geographic separation of these isolates. Evolutionarily this implies that the encoded ladderane could potentially be ancestral, resulting in a subsequent loss by the majority of mainland and Chatham Island isolate.

Genome searches for differences related to plasmid acquisition showed that no definitive genes on the chromosome could be positively or negatively correlated to the presence or absence of either pADAP mega-plasmids or other plasmid types. Bacterial plasmids overall adapt to shift specificity to the host bacteria as an element of long-term persistence in the population, increasing stability. Several mechanisms for coevolution of bacterial host and plasmid have been shown by Sota et al. (2010), where point mutations on the plasmid replication gene were demonstrated to improve stability in the host range. Further studies show that in addition to mutations in the plasmid transcription regulation gene, mutations in the regulator *fur* gene (Stalder et al. 2017).

Chromosomal searches for extracellular lipases confirm what was observed in Chapter 4 section 4.2.5.2. One extracellular phospholipase (A1) is encoded on each *S. entomophila* genome. Extracellular phospholipase A1 has previously been implicated as a secreted lipase in *S. ficaria,* as a growth phase-dependent secretin (Givskov et al. 1988). The secretion pathway of PhlA in *S. marcescens* has been described as flagella-dependent, relying on *flhDC* regulator. Coregulation of flagellar and phospholipase was implicated in a study undertaken by Anderson et al. (2017), but the role of PhlA as a virulence factor could not be determined. Due to the consistent response across both pathogenic and non-pathogenic isolates in plate assays found in this study (Chapter 4, section 4.2.5.2), this again cannot be established as a significant difference between pathogenic and non-pathogenic isolates, although there is the potential that this phospholipase can indeed aid host colonisation. A trend was noted in Chapter 4 in both protease and lipase response plate assays that higher production was seen in the pathogenic *S. entomophila* isolates. Encoded by *S. entomophila* isolate 626, 126 peptidases and

9 lipase-associated genes were identified including 3 phospholipase. Previous studies have shown an association with a pathogen's ability to cleave phospholipids in eukaryotic cell membranes, therefore mediating colonization of a host. Secretion of phospholipase B such as the orthologue encoded on the chromosome of *S. entomophila* isolate 626, has previously been associated with increased virulence in *Candida albicans*, a fungal pathogen of humans (Leidich et al. 1998).

The presence of chromosomally encoded lipases, chitinases, and proteases suggests that some functional basis of pathogenicity is controlled by the chromosome and not solely the plasmid- though these relate mostly to colonization, adhesion, and degradation of the insect hindgut. These accessory enzymes are mostly uniform across the chromosomes of *S. entomophila* isolates, showing conservation in the accessory virulence factors encoded. AntiSMASH detected seven secondary metabolite clusters located on the *S. entomophila* 626 chromosome, including three with BLAST results showing similarity to previously described metabolites. This includes a siderophore with 70% identity to aerobactin identified previous from *Xenorhabdus szentirmaii,* a *tycC* gene, an antibiotic gene cluster postulated to affect sporulation in the soil-dwelling bacterium *Bacillus (=Brevibacillus) brevis* (however, lacks the appropriate mechanisms for peptide synthase from TycA and TycB), and enterobactin. Enterobactin is utilised by gram-negative bacteria to chelate iron from the environment or from a host, such as demonstrated in *Klebsiella pneumoniae* (Bailey et al. 2016). Presumably, this has an iron-chelating functionality that allows *S. entomophila* to directly sequester iron compounds from the host. Various other peptide antibiotic genes associated with gram-negative bacterium were identified including a beta lactone, hserlactone, thiopeptide, and a type 1 polyketide synthesis gene. A putative terpene cluster was predicted in ten isolates of *S. entomophila*. The fact that these regions are not consistent across all isolates of *S. entomophila* could provide evidence as to whether the loss is occurring with large redundant gene clusters in pathogenic isolates, although the encoded terpene is not unique to non-pathogenic isolates. That this terpene encodes a similar cluster to sodorifen, which responds to fungal volatiles in the soil, could be an indicator that the genome of *S. entomophila* is specializing towards a potential host-dependent. Production of sodorifen by *Serratia plymuthica* where, although the function has yet to be derived (Domik et al. 2016), regulation is tightly linked to the carbon catabolite repression system (Magnus et al. 2017). Although fungal volatile compounds are known to be communication signals, it is not yet known how or if bacteria can utilise or recognize these signals (Schmidt et al. 2015). If soil bacteria could utilise these compounds in the soil as a communication or defence mechanism, a bacterial response could benefit a free-living organism, especially in grass grub-free regions.

Amongst the unique genes found in *S. entomophila* was the 3 gene itaconate cluster *mcl*, *hcaR,* and *mcH* involved in pyruvate and acetyl-CoA assimilation via ethylmalonyl-CoA pathway. Itaconate

medium has previously been used as a marker to differentiate *S. entomophila* from other *Serratia* spp. on media (O'Callaghan 1993). Itaconate is now described as a metabolite that inhibits pathogenic bacteria from cleaving isocitrate into succinate in the glyoxylate cycle, which has been implicated in pathogen persistence in fungi and bacteria (Chew et al. 2019). A pathogenic bacterium's ability to degrade itaconate would be a beneficial trait for out-competing other bacteria. Notably, this metabolite cluster has been previously described in *Y. pestis*, and allows the bacteria to convert itaconate into pyruvate and acetyl-CoA, allowing the pathogen to utilise a host defense as a metabolic process (Sasikaran et al. 2014). Two hypotheses were developed as to why it is advantageous for *S. entomophila* to acquire the itaconate degradation operon. Initially, it was thought that if the presence of this cluster in *Y. pestis* allowed for survival in a mammalian host- adaptations might have occurred that would benefit *S. entomophila* in an insect host, therefore allowing the bacteria to survive long term. A second hypothesis was developed after more literature research that the itaconate degradation pathway could be taking the role of another energy metabolizing pathway, potentially using fungal secreted itaconate substrate as a carbon source (van der Straat et al. 2014). Saprophytic fungi are often seen in the cadavers of amber disease-affected larvae, therefore the link to fungi can't be dismissed and warrants further investigation. Relating to COG classifications in section 5.2.1, differences between *S. entomophila* isolate 626 and the type strain isolate for *S. ficaria* in encoded metabolic pathways for carbohydrate metabolism could potentially be linked to the utilisation of other pathways via trophic specialisation which is to be discussed more in-depth later.

Although not fully determined, it can be assumed that the functionality of the itaconate operon in *S. entomophila* is unlikely to be related to survival in host hemolymph protozoa. More research could be undertaken to validate this theory. The lack of sterile, pure sources of protozoan cultures hindered attempts to test this hypothesis. Targeted mutagenesis of the *mcl* gene interrupting the itaconate degradation pathway revealed that in competition assays the wildtype isolate bears a competitive advantage against the mutant Mcl1::tet in M9 minimal media and *in vivo* grass grub bioassays. It was determined that the itaconate degradation pathway in *S. entomophila* is likely functional in the utilization of itaconate as a potential carbon source as demonstrated through the inability of the Mcl1::tet mutant to grow on itaconate media. The comparison of the *mcl* mutated 626 isolate to the wildtype isolate shows that the itaconate cluster conveys some level of metabolic advantage to. It can be assumed that this pathway takes the place of a necessary energy source that is beneficial for *S. entomophila*. Further comparisons will be undertaken (Chapter 6) to determine any niche pathway that the itaconate degradation cluster could occupy in comparison with other members of the *Serratia* genus. Potentially, the acquisition of this cluster has mediated the loss of another pathway to bypass energy costing routes to generate pyruvate (synthesized from the itaconate pathway). Acquisition of this region is hypothesized to have been recently in the evolution of *S. entomophila* as a species, based

on the G+C content of the region being dissimilar to that of the wider chromosome and the inversion gene operon orientation between difference sequenced isolates of *S. entomophila*. Comparing the orientation of this operon in specific isolates does not reveal any link between inversion and MLSA phylogeny clade groupings.

When specifically looking at regions of horizontally acquired or phage integrated DNA, historic acquisition events are usually characterised by amelioration of the introduced DNA to reflect the G+C content of the chromosome over time (Lawrence and Ochman 1997). Evolutionary pressure could have allowed *S. entomophila* to lose this unknown metabolic module after acquiring this cluster that could allow for utilization of a secondary product from soil-borne fungi.

In this analysis, there was no indication between any isolates that there were any additional beneficial factors on the chromosomes of *S. entomophila* isolates that are pADAP bearing compared to non-pathogenic conspecifics. These isolates may not have encountered pADAP in their ecological niche or have sustained a population devoid of pADAP. In the case of *S. entomophila*, some isolates are non-pADAP bearing but harbour an alternate plasmid, for example, the Chatham Island isolates 440, 442, and 294. Plasmid types that contain a similar origin of replication are known to be incompatible with one another forming incompatibility groups. Velappan et al. (2007) showed that even when stability time can be increased in a laboratory-based environment, plasmid loss is eventually observed. A study by Grkovic et al. (1995) found that in forced conjugation of *S. liquefaciens* isolate 377 (incorrectly classified as *S. proteamaculans*), the original plasmid was then lost in favour of the recipient plasmid, supporting that these plasmids have incompatibility types. Interestingly, we see in the assessed isolates of *S. entomophila* that even geographically and genomically similar isolates such as 626 (pathogenic, plasmid bearing) and 477 (non-pathogenic, plasmid free) show that chromosomal encoded factors are not the only prerequisite to harbouring pADAP. It has been previously documented in studies that the ability for biofilm formation produces a suitable environment for the exchange of conjugative elements such as plasmids. In *S. marcescens*, this biofilm formation is mediated by *bsmA* and *bsmB*, quorum sensing mediated genes crucial for the production of effective biofilms (Rice et al. 2005). This operon is present across all isolates of *S. entomophila*, but differences in the regulatory mechanism via quorum sensing could mediate different outcomes for different isolates, either enabling or preventing efficient biofilm production. Conversely, there could potentially be a chromosomally-bound regulatory system that aids in the incorporation of the plasmid into the host cell. This will be explored in Chapter 7, where the ability to swap plasmids between varying chromosomal backgrounds will be investigated and discussed.

Considering the *in silico* and *in vitro* results presented in this Chapter and Chapter 4, it is probable that *S. entomophila* is not currently undergoing strict genome reduction, though comparisons with other species such as *S. proteamaculans* may show whether restrictions are occurring (discussed further in Chapter 6). *S. entomophila* does, however, have unique regions that could allow for niche colonisation and persistence, namely the itaconate degradation cluster. Further analysis would be needed to elucidate any regulatory factors that could contribute to the ability of a chromosome to persist with pADAP. Gene-level studies however do not indicate that anything lies within the chromosome that would provide a beneficial or metabolic advantage for a specific isolate to encode for pADAP whilst others remain plasmid-free.

Chapter 6

Comparative genomic analysis of isolates of *Serratia entomophila* and *Serratia proteamaculans*

6.1 Introduction

In comparison to *Serratia entomophila,* analysed in Chapter 5, *Serratia proteamaculans* display a wider variety of disease phenotypes and a broader range of plasmid types acquired. To date, there has not been any exploration of chromosomal differences that would allow *S. proteamaculans* to favour any specific environmental niche, or whether this bacterium is evolving independently as a highly virulent entomopathogen.

The availability of complete bacterial genomes has exponentially increased through rapid development of next-generation sequencing. Combined with decreasing sequencing run costs and affordability of long-read data either by Illumina or PacBio and, more recently, nanopore sequencing, critical comparison of isolates is possible. In a concurrent project, plasmid differentiation in amber disease-causing *Serratia* was characterised by Sitter (2020). Chromosomal differences however were yet to be defined. By comparing large genome datasets of chromosomal sequences of *S. entomophila* and *S. proteamaculans* isolates, genetic associations with pathotype can be defined. This could suggest what large-scale differences could account for the variability of pathogenicity seen within *S. proteamaculans.* Identification of genomic islands and transposable elements on the chromosome may explain why there are larger regions of variability in *S. proteamaculans*, and what link this has with pathotype. Identifying genomic rearrangements and associating them with pathogenicity may help define the benefits these organisms have in reduction and/or large-scale heterogeneity.

Here, a comparative analysis is undertaken using genomes developed for 56 isolates were utilised in total comprising 23 *S. entomophila*, 28 *S. proteamaculans* in addition to 5 further Yersiniaceae as comparisons. These isolates were of varying pathotypes, characterised mainly as pathogenic or non-pathogenic. Six isolates of pathogenic *S. proteamaculans*, AGR96X, LC, 20093, 10Novel, 1129, and MH5, were further characterised as hypervirulent, due to their ability to cause rapid death of the grass grub. Most of these encode an AfpX variant of the antifeeding prophage on their plasmids. The isolate 10novel however, does not encode this Afp variant and instead encodes a TC cluster on its plasmid, while still causing hypervirulence. Therefore, in addition to a general comparative analysis of the two

amber disease-causing species, the chromosome of various *S. proteamaculans* will be explored to determine what degree of dissimilarity is observed among pathotypes and in contrast to the wider *Serratia* group.

6.2 Results

6.2.1 Composition and comparison of *S. proteamaculans* genomes

Twenty-eight isolates of *S. proteamaculans* were used in this analysis. Sequence comparisons of these isolates revealed a diverse range of encoded tRNAs, rRNAs, and coding sequences (Table 24), implying there are large regions of variations between isolates. Of the isolates with plasmids encoding the anti-feeding variant AfpX, five have a chromosomal G+C content of ~55.1%, similar in content to that of chronic and non-pathogenic (standard Afp carrying plasmids) isolates. The TC encoding 10novel isolate, however, has a lower chromosomal G+C content at 52.1%, having the lowest chromosomal G+C content of any sequenced *S. proteamaculans*.

The G+C content of *S. proteamaculans* isolates, on average, deviates from that of *S. entomophila* by ~4% (Figure 50). Wilcoxon rank-sum P-value test determined there was a significant difference between the values of these two species G+C% ($p<0.001$). When compared to that of *S. entomophila* (Chapter 4), the mean number of CDS encoded on the chromosomes is higher as well as the overall bp length of the chromosome. A larger range of rRNAs and tRNAs are also encoded in *S. proteamaculans*, showing a higher range of diversity across these chromosomes than in *S. entomophila* sequences.

The non-pathogenic isolate 3041 has the highest number of CDS of the *S. proteamaculans* analysed, at 5,171. The lowest was isolate 12a with 4,439, revealing a broad range difference of 732 CDS across isolates (Table 24). This range of encoded CDS is 1.4× that of which was observed between *S. entomophila* isolates. Within the *S. proteamaculans* chromosomes, those encoding AfpX plasmid variant had the smallest range of encoded CDS where the difference between highest encoded (MH5) and lowest (AGR96X) is 195. Overall, the encoded rRNA operons are variable. Hypervirulent isolates AGR96X and MH5 show the lowest number of predicted encoded rRNAs covering one full operon and an orphan 5s rRNA. Although this is not consistent, other AfpX encoding *S. proteamaculans* again have lower rRNA operons: isolate 20093 has 6 operons, 3 with 3x rRNA and 3 with singular rRNA genes, whereas others have 5 (2 x 3 genes encoded and 3 x 1 gene). Variability is also seen within the other pathotypes, with no consistent trend being observed across chronic or non-pathogenic isolates. Only 6 of the 33 (18%) isolates have high numbers (18-22 individual associated genes) of encoded rRNAs as opposed to the 30% of *S. entomophila.*

Between the subspecies of *S. proteamaculans* described in Chapter 3, little difference in overall genome statistics is observed. The G+C content is relatively conserved as a whole, with the range of CDS encoded on the genome being small. No subspecies split is observed in the number of rRNA encoding operons or the number of tRNAs present on the chromosome.

Table 24 Genome statistics of individual *Serratia proteamaculans* isolates

Isolates with a bolded name are isolates under the subspecies *quinovora* as defined in Chapter 3.

	S. proteamaculans isolate								
Feature	4[1]	142[1]	143*	145[1]	149[1]	299[1]	336[1]	465[1]	1048[1]
Size (bp)	5,261,985	5,211,076	5,441,365	5,433,271	5,097,608	5,494,205	5,427,128	5,311,017	5,439089
G+C content (%)	55.1	55.2	55.1	55.2	55.1	55.0	55.0	55.1	55.0
CDS	4,846	4,863	4,998	4,967	4,855	5,093	5,024	4,892	5,111
tRNAs	85	87	85	78	75	76	78	78	83
rRNAs	13	22	22	12	16	10	10	14	22

	S. proteamaculans isolate									
Feature	1071[1]	1129[2]	1137*	1457[1]	1769[1]	**3041***	20093[1]	12a*	12newD*	25e*
Size (bp)	5,401,630	5,411,428	5,604,040	5,374,788	5,400,292	5,458,427	5,431,358	5,287,754	5,241,880	5,304,470
G+C content (%)	56.4	55.1	56.2	55.0	55.2	55.2	55.0	55.1	55.2	55.1
CDS	4,981	4,999	5,100	4,976	4,971	5,171	5,006	4,439	4,807	4,880
tRNAs	80	77	83	75	82	83	84	84	78	80
rRNAs	12	10	13	9	11	15	12	12	12	10

	S. proteamaculans isolate								
Feature	**AGR96X**[2]	CfB*	G*	K*	LC*	M*	MH5[2]	Novel10[2]	Puna18*
Size (bp)	5,497,482	5,427,745	5,352,346	5,388,719	5,435,284	5,347,286	5,524,696	5,304,470	5,496,333
G+C content (%)	55.1	55.2	55.0	54.9	55.1	55.0	54.9	55.3	55.1
CDS	4,977	5,000	4,973	5,000	5,048	4,921	5,172	4,877	5,042
tRNAs	81	79	83	79	81	79	72	80	91
rRNAs	4	7	13	9	7	13	3	11	23

[1] Isolates encoding pADAP
[2] Isolates encoding AfpX
*Plasmid free isolates

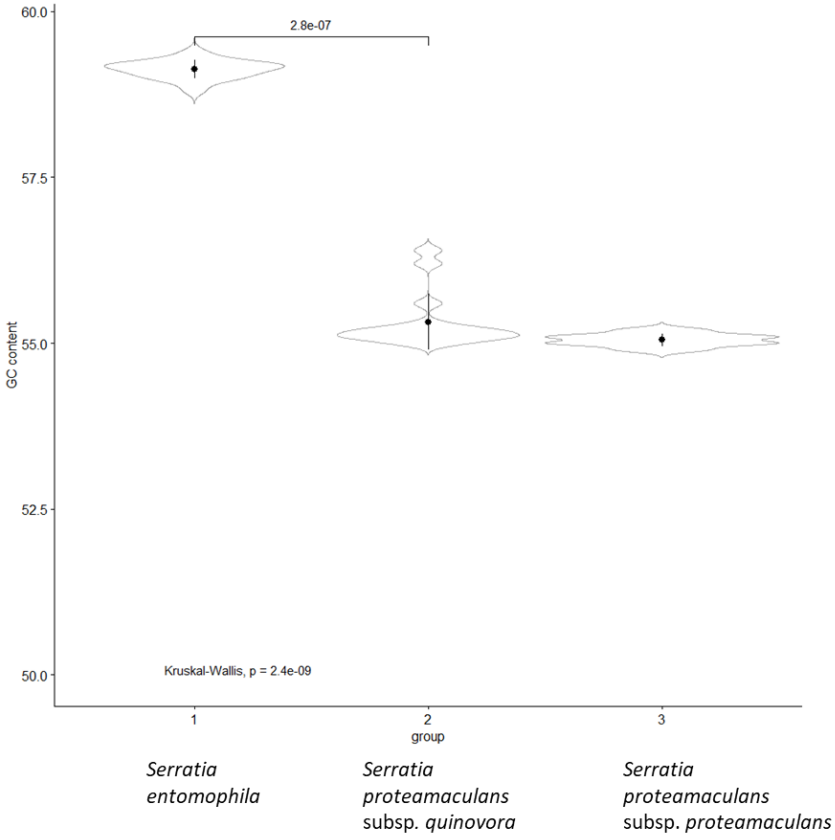


Figure 50 Distribution of genomic G+C content of isolates included in this study
P-value calculated by Kruskal Wallis P-value

Average nucleotide identity (ANI) was calculated between isolates of *S. proteamaculans* at the species level, before including isolates of *Serratia* spp. to confirm grouping of isolates amongst species. Nucleotide identity was more varied amongst *S. proteamaculans* (Figure 51) than seen in *S. entomophila* (Chapter 5), which was expected given there are two known subspecies of *S. proteamaculans* (subsp. *proteamaculan*s including LC and 149, and subspecies *quinovora* including AGR96X and 3041).

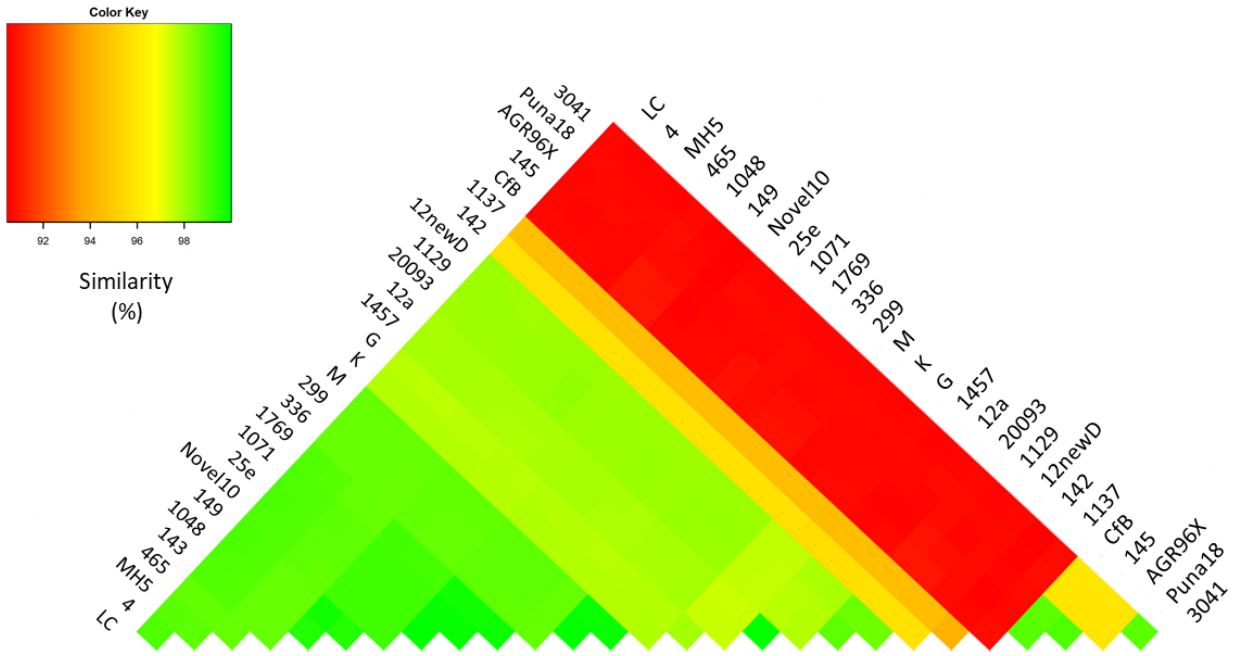


Figure 51 Average nucleotide identity of *Serratia proteamaculans* isolates (%)
The percentage is shown in the colour key, where yellow indicates the highest degree of percentage similarity.

For large-scale comparative analysis, *S. entomophila* isolates listed in Chapter 4 were incorporated in addition to *Serratia* spp., including *S. liquefaciens, S. ficaria,* and *S. marcescens*, and the *Yersinia frederiksenii* isolate Y49 (Table 25). Overall, the chromosomal characteristics of these isolates did not differ greatly from isolates of *S. entomophila* and *S. proteamaculans*. G+C% content is for the most part in line with what is observed in *S. proteamaculans* (Table 24). *S. ficaria* shows a high degree of similarity to that recorded in *S. entomophila* (Chapter 5). This similarity is shown further in Figure 52, with ANI of the *Serratia* spp. showing high degrees of similarity between *S. entomophila* and *S. ficaria*. Chromosomal size and the number of encoded CDS is larger than that of *S. entomophila* for this species but is similar to *S. proteamaculans*. Overall, one of the largest differences is the lower rRNAs encoded in the extended *Serratia* family than in both isolates of *S. proteamaculans* (Table 24) and *S. entomophila* (Chapter 5, section 5.2.2). Due to there being fewer isolates for each of these outgroup species and that the high numbers of rRNA operons in *S. proteamaculans* and *S. entomophila* is not uniform, we cannot conclude that this is uniform for the species.

Table 25 Genome statistics for additional *Serratia* spp. and *Yersinia* used in this study

		Yersiniaceae			
		S. liquefaciens		*S. marcescens*	*Y. frederiksenii*
Feature	F28	377	376	1577	Y49
Size (bp)	5,427,745	5,287,990	5,290,681	5,435,284	5,105,482
G+C content (%)	55.2	55.5	55.5	55.1	47.2
CDS	5,000	4,836	4,843	5,048	4,717
tRNAs	79	82	79	83	79
rRNAs	7	10	9	11	22

Including other members of the *Serratia* spp., it is easier to see the species clustering based on their nucleotide identity (Figure 52). *S. entomophila* isolates fit close together with a shared identity above 95% amongst all isolates. For the most part, clustering is also observed with *S. proteamaculans* isolates. The main difference is the distinction of subspecies in *S. proteamaculans, S. proteamaculans* subsp. *proteamaculans* and *S. proteamaculans* subsp. *quinovora*. AGR96X is already identified as subspecies *quinovora*, as corroborated by 16S phylogeny (Hurst et al. 2018). Alignment, therefore, shows that CfB, 145, Puna18, and 3041 also fit into this subspecies, whereas the large majority of the remaining isolates group together into subspecies *proteamaculans*. *S. marcescens* isolates (1577 and 457) also show a closer ANI similarity to *S. entomophila* than with *S. proteamaculans.*

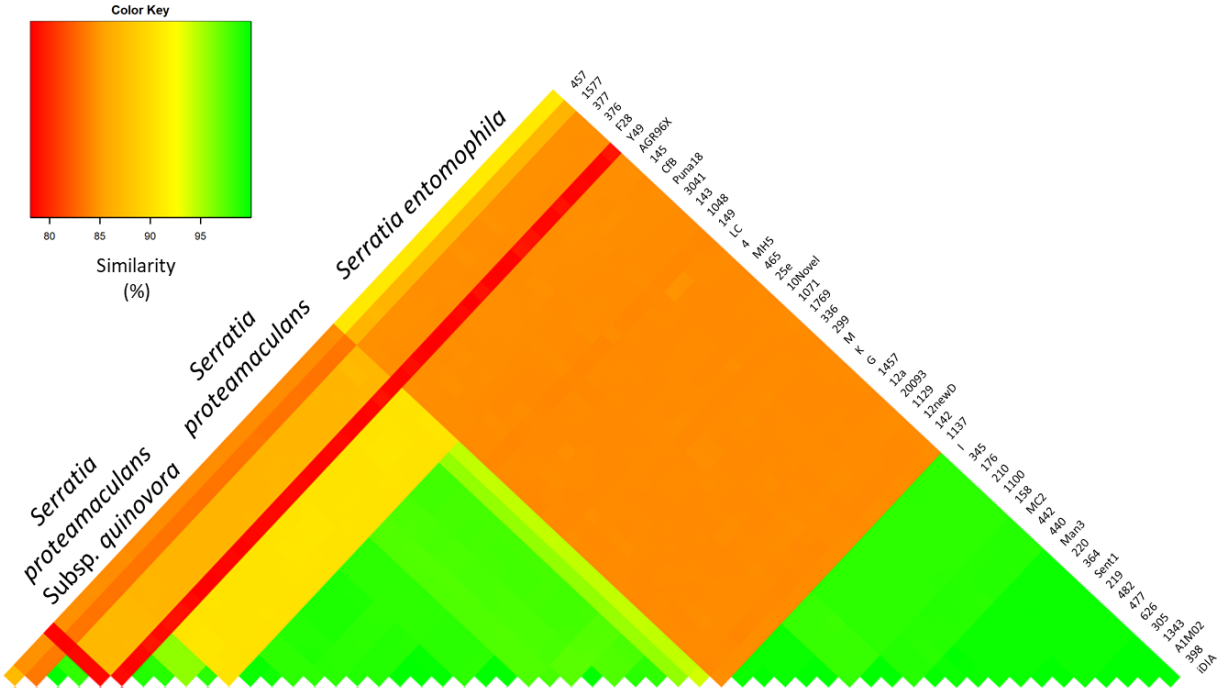


Figure 52 ANI values of all isolates assessed in this study
Isolates clustered to similar ANI values. Species shown on the left with individual isolate identity shown on the right. Percentage represented as a colour change. Isolate Y49 refers to an isolate of *Yersinia frederiksenii.*

6.2.2 Full core phylogenies of *S. entomophila* and *S. proteamaculans*

As discussed in the previous Chapter, 16S rRNA is a good indicator of the clade orientation of many species. However, with closely related isolates within the same species, the benefits of 16S are lowered, and limitations arise. Based on the hypothesis that *S. proteamaculans* isolates have high levels of gene acquisition from the environment, it would be appropriate to utilise a method that mitigates the effects of this on phylogenetic output. Core genome phylogenies of both species were therefore constructed using Gubbins. Inputting a core SNP alignment allows Gubbins to identify regions of elevated variation resulting in the exclusion of these from analyses and therefore minimising the effects of horizontal gene transfer on phylogeny. In this way, a more accurate evolutionary tree based on the core genome can be generated.

From this core phylogeny (Figure 53) the lineage leading to *S. entomophila* has undergone less change than isolates of *S. proteamaculans* since it split from a predicted common ancestor. Isolates of *S. proteamaculans* have diverged into separate subgroups, although closely related. We can assume, based on Figure 53, that variation between the subspecies including AGR96X (subsp. *quinovora*) is broader, as more change can be observed intraspecies than within the clade including MH5 and 10Novel.

Core genome phylogeny splits the cohort of *S. entomophila* isolates into three; one with the previously known differences of the Chatham Isolates (440, 442, 294 and West Coast isolate MC2), but with further differentiation. Isolates I, 158 176, 345, 1100, and 210 group together separately from the main cluster of *S. entomophila* isolates. This will be elaborated on in more detail later in the analysis, but it can be assumed that since the % ANI is close between these isolates and the main cohort *of S. entomophila*, the presence of an additional cluster of genes acquired could be responsible for the separation.

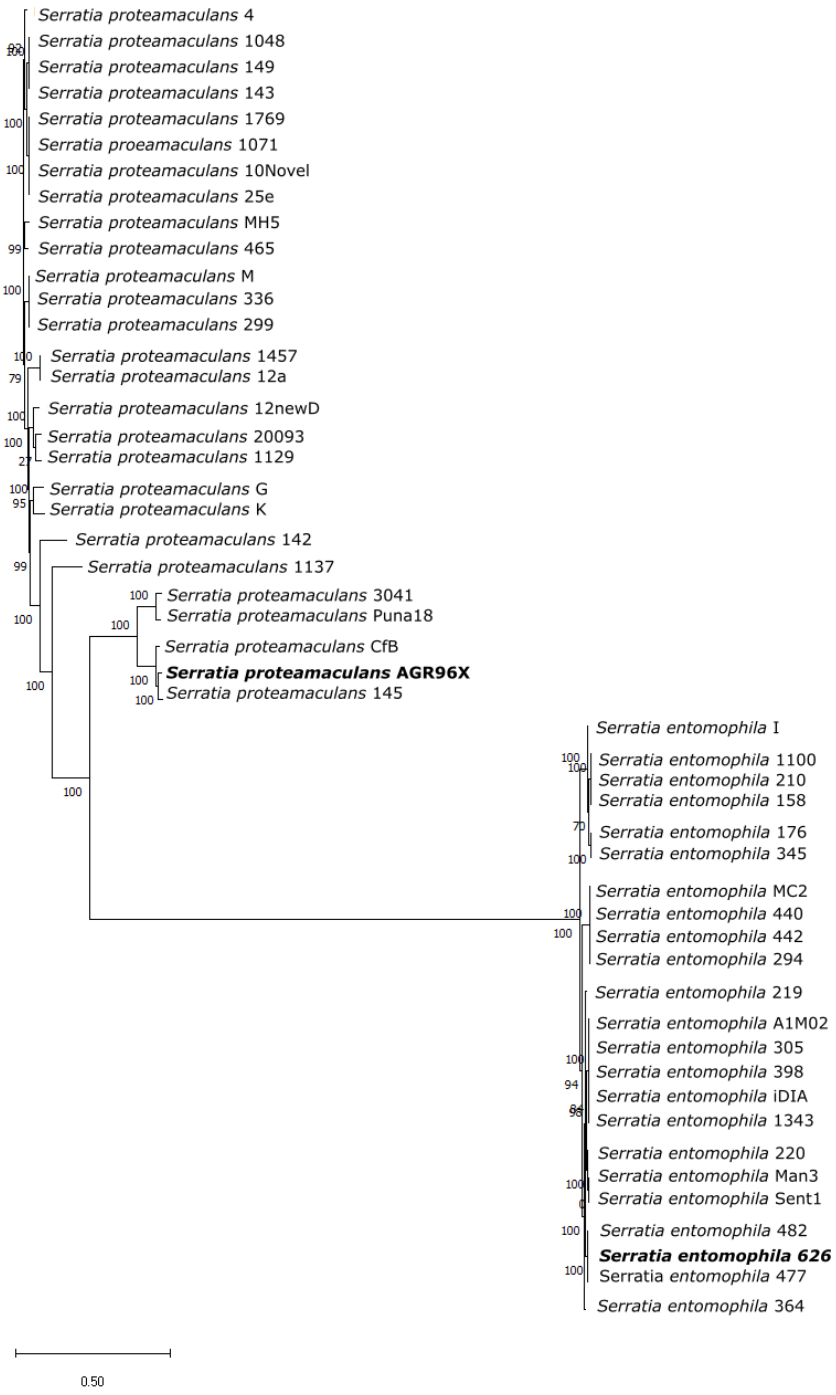


Figure 53 Core genome phylogeny of closest related *Serratia* isolates
Core genome calculated using Snippy and corrected for recombination events using Gubbins. Maximum likelihood tree calculated using FastTree. Support values (proportion of 1000 trees) were generated by resampling the data with 1000 replicates. The scale bar indicates branch length in the number of substitutions per site. Bold denotes *S. entomophila* isolate 626 used in a current biopesticide and hypervirulent isolate of *S. proteamaculans* AGR96X.

6.2.3 BRIGs comparison analysis

BRIG diagrams of sequences aligned to a central reference for each of *S. entomophila* and *S. proteamaculans* are shown in Figure 54. As both are large images with many chromosomes displayed the individual key references for isolates were not able to be included on a visible image and are thus included in Appendix D.1. The chromosomes of *S. entomophila* isolates 442 and 440 (outer rings) from the Chatham Islands and MC2/219 isolates (3rd and 4th outer rings) from the West Coast of the South Island show a larger degree of chromosomal variation than other isolates (Figure 54A). *S. entomophila* is more homogenous across chromosomes than observed in *S. proteamaculans,* whereby regions of differentiation are similarly located across all isolates (54B). No visible regions of differentiation can be observed between pathogenic and non-pathogenic isolates (warmer and cooler-toned rings). Annotated regions of difference to the reference chromosome allude to areas of *S. entomophila* where change is observed (Figure 54). Region 1 encodes for an area of high variability across *S. entomophila* genomes, predicted as a genomic island (see this Chapter, section 6.2.6). Similar is observed for region 2, with the reference 626 chromosome region encoding a prophage integrase flanked region with putative fimbrial genes encoded between. In 626, region 3 encodes 44 genes including a transposase. This area of the chromosome is packed with hypotheticals, two HTH-type transcriptional regulators, and some transporter-associated genes. In the Chatham Island isolates this variable region is approximately 100 Kb long as opposed to the region in 626 at approximately 55 Kb. A 13 Kb region (region 4) follows the trend of the previous three regions where variability is seen across all the chromosomes included (section 6.2.6). Again, this area encodes a large number of hypothetical proteins, though isolate 626 encodes two Type I restriction enzymes, as well as *hipA*, a serine/threonine-protein kinase toxin. This variable region is not present in isolates from the Chatham Islands or in West Coast isolate MC2, indicating they were not exposed to this HGT event.

Where there are regions of differentiation, they are largely in similar locations to those seen in *S. proteamaculans*. The *S. proteamaculans* chromosomes however have many more regions that are not aligned to the reference, which is to be expected as the 568 isolate is not of New Zealand origin. Originally this isolate was selected as a reference genome due to the known completeness of the genome. Further BRIG alignments utilising AGR96X hypervirulent isolate as the reference was also conducted as a comprehensive reference of New Zealand isolates (Figure 55). These seem to be generally consistent across most isolates with no region-specific to a recognized phenotype enabling the differentiation between hypervirulent, chronic, and non-pathogenic isolates. There seem to be no obvious groupings within *S. proteamaculans* whether the isolate is pathogenic or non-pathogenic. A large region at 4.5 Mb (indicated as 2 in Figure 54B), similar to what was seen in *S. entomophila,* encodes a region of variability across all isolates. This area has a high density of fibrin-associated

proteins, as well as a multidrug resistance (*mdtO*) protein in isolate 143. Region 1 that isolates of *S. proteamaculans* have consistent differences to *S. proteamaculans* isolated outside of New Zealand. The same can be seen for region 3. Region 4, however, is present in all isolates with varying encoded genes. Mauve alignments (section 6.2.5) show this area as a standalone collinear block of varying sizes across isolates. Again, this area is filled with hypothetical proteins and fimbrial proteins in addition to the presence of several transposases. Exploration of the chromosomes further in sections 6.2.5 and 6.2.6 implicates a lot of these regions to HGT events, showing the bounds of the pangenome increases with exposure to foreign genetic material.

As BRIG uses a reference genome to map other strains, this may not be the most comprehensive comparison as this merely shows differences from the reference genome. For this reason, reference genomes were selected based on their applicability to the study and size relative to the extended cohort of isolates, therefore isolate 626 for *S. entomophila* and isolate AGR96X for *S. proteamaculans* were selected. The Chatham Island *S. entomophila* isolate 440, with the largest of all the *S. entomophila* chromosomes, was also used as a reference genome. It was also pertinent to include this comparison to determine whether there were any regions lost in other isolates that are present in Chatham Island isolates or West Coast isolate MC2. Through BRIG analysis any regions present on a chromosome that isn't in the reference genome are not represented, and therefore only display genomic omissions rather than additions, although BRIG visuals are simpler to interpret data on a large scale. Due to this, Mauve was utilised in addition to BRIG comparisons to compare key isolates and find variations between a reference and query genome.

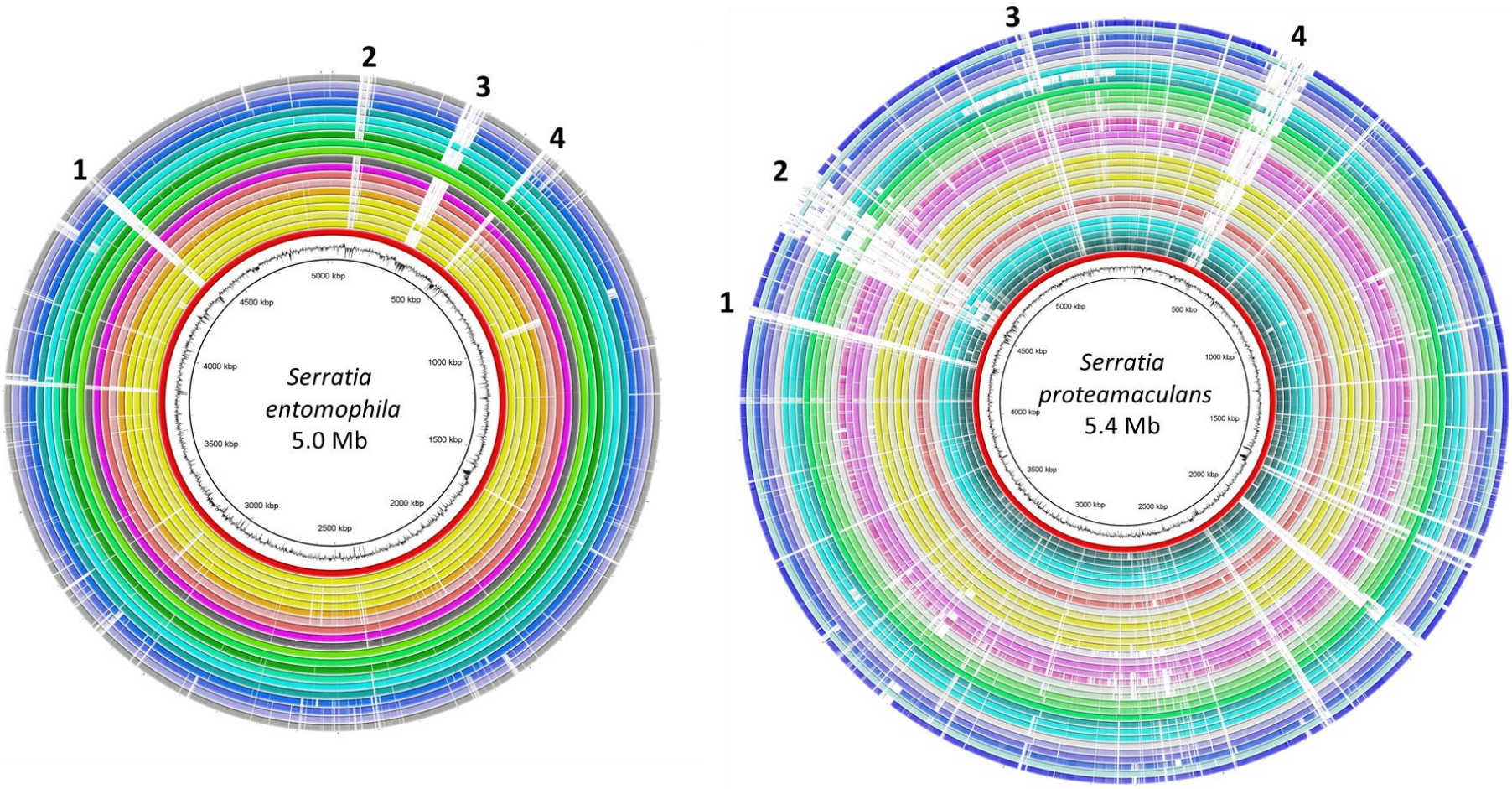


Figure 54 BRIG diagrams of both *Serratia entomophila* and *Serratia proteamaculans*.
A) *S. entomophila* 626 as the reference genome (shown in red) with rings in 'warm' colours e.g., Yellow orange pink indicative of the chromosome of pathogenic isolates. Outer 10 rings are non-pathogenic isolates with the 2 grey outside rings being the isolates from the Chatham Islands, and the 2 in from that being West Coast isolates. B) *S. proteamaculans* GenBank sequence 568 as the reference genome (shown in red). Colour scheme follows that of *S. entomophila* with warm colours indicating pathogenic, and cooler colours being non-pathogenic isolates. Inner 5 green rings are 'hyper-pathogenic' isolates. Regions highlighted as numbers are explored in the main text.

BRIG does show, however, the unique regions on hypervirulent isolate AGR96X (innermost green ring on Figure 54B). Genome searches identified the specific genes in these regions (Figure 54), with further analysis in section 6.2.6 corresponding some of these regions to putative genomic islands. Identification of potentially unique regions on the chromosome that could allude to the evolution of a hypervirulent pathotype could be explored. Figure 55 shows that the regions of differentiation in the broader *S. proteamaculans* cohort in relation to isolate AGR96X follow the same regional pattern as seen in Figure 54B. Regions of difference also align with predicted genomic islands (section 6.2.6.1). This result implies that even within diverse *S. proteamaculans* (such as non-New Zealand origin isolate 568) regions of differentiation could tentatively be assigned to HGT events. Regions highlighted in Figure 55, indicate differences that could potentially be beneficial to AGR96X. At region one, a small 5 Kb predicted region of difference was shown between AGR96X and other *S. proteamaculans* isolates. This corresponds to a genomic island (GI) identified in section 6.2.6.1. The same is noted for regions three, four, and five where putative GI's correspond to regions of reduced homogeneity. However, this is not observed in region two, implying that other mechanisms of chromosomal changes such as inversions or mutations have influenced the changes seen here.

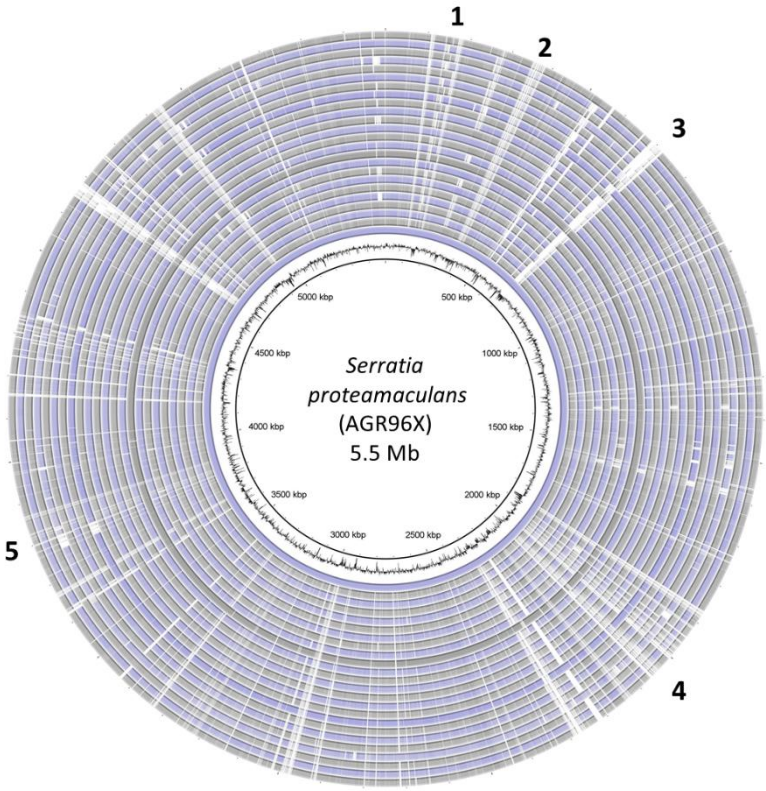


Figure 55 Alignment of hypervirulent AGR96X to other isolates of *Serratia proteamaculans*
Numbers indicate regions of difference between AGR96X and other isolates of *S. protemaculans*. Regions of difference shown as white gaps between homoegenous sequence and align with predicted genomic islands of AGR96X shown in Figure 62 and regions of difference in Figure 54B.

Aligning the additional *S. entomophila* chromosomes against the larger Chatham Island isolates reveals that large regions of disparity of the North and South Island isolates when compared to the Chatham Islands and West Coast isolate MC2 (Figure 56). The large region at 1.2 Mb (indicated with number 3 on Figure 56) aligns with what was observed in Mauve alignments in Chapter 4. Large regions of differentiation align with the displayed putative genomic islands also shown in Chapter 4.

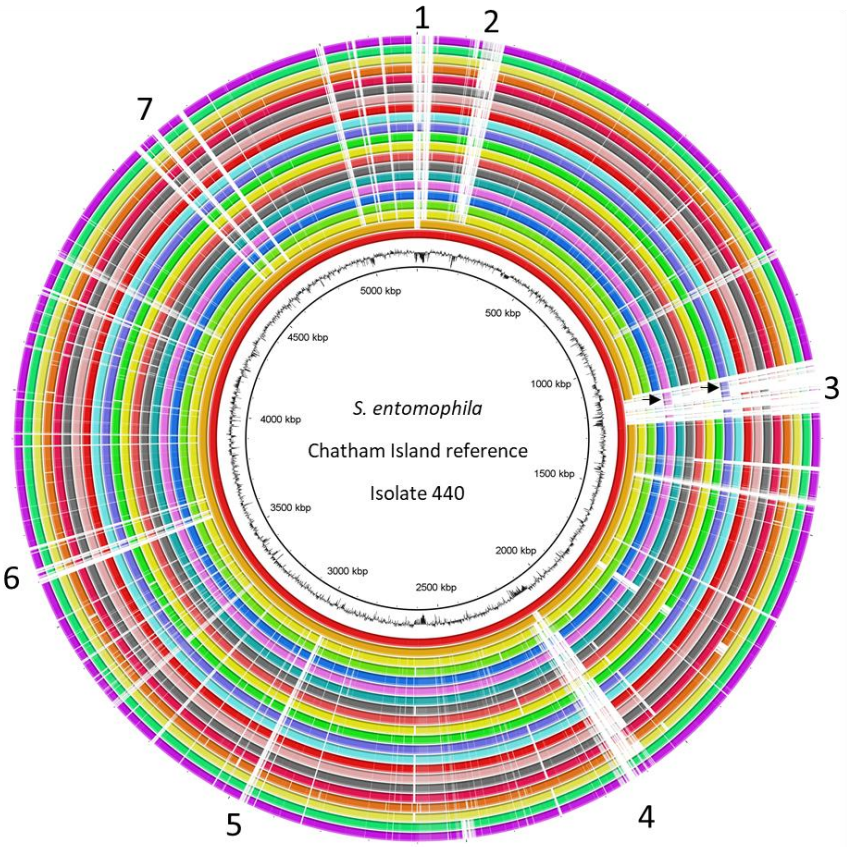


Figure 56 Alignment of *Serratia entomophila* isolates to Chatham Island isolate 440
Innermost rings from additional Chatham Island isolates (442 and 294) and West Coast isolate MC2. Shows difference of additional regions in these larger chromosomes than that of *S. entomophila* 626. Further detail is found in section 6.2.6. Full key of individual isolates correlated to which ring can be found in Appendix D.1. Black arrows indicate two isolates with more similar regions at 1.25 Mb (isolates 176, pink and 364, purple rings). Numbers indicate regions of difference between the Chatham Islands and other isolates that correspond to putative genomic islands (Chapter 4).

6.2.4 Determining the presence of advantageous metabolic pathways

As was previously discussed in Chapter 5, *S. entomophila* in comparison to reference *S. proteamaculans* 336X showed a reduced number of COG-affiliated genes with metabolism. To determine whether *S. proteamaculans* had an increased number of metabolically encoded pathways that could be beneficial for bacterial survival, KofamKOALA profiles of representative pathogenic and non-pathogenic isolates of each species were undertaken. Determining any interspecies and hypervirulent specific pathways could then aid in the metabolic fitness of an isolate. From the resulting protein annotations, KEGG Mapper was then implemented to identify pathways mapped using the KEGG PATHWAY database, identifying high-level functions present. By comparing these missing or

additional functionally complete pathway modules were identified between each species. A selection of isolates representative of putative pathotype, species, and geographic isolation and their number of encoded metabolic modules is displayed in Table 26. Full results for all isolates are shown in Appendix D.2

KEGG outputs were inspected for any major differences between complete metabolic pathways. Across all isolates assessed, predictions for *S. proteamaculans* isolates have one extra complete carbohydrate metabolism system than those of *S. entomophila*. D-galactonate degradation system via the Leloir pathway is uniform across all isolates, but *S. proteamaculans* also encodes the De Ley-Doudoroff pathway. This system involves seven proteins, including the sub-pathway *dgoD*, *dgoK,* and *dgoA* genes that degrade D-galactonate to D-glyceraldehyde 3-phosphate (G3P) and pyruvate. Pyruvate is then synthesised from G3P in the glycolysis pathway (*gapA, pgk, gbmA, eno, pykF*). This pathway is listed as incomplete in *S. entomophila*, with only two of the pathways required genes present on the chromosome. Additional pathways for degradation of D-galactonate could allow *S. proteamaculans* to have a metabolic advantage, utilising further methods for energy sequestering. This result is consistent with what was observed via COG analysis of *S. entomophila* 626 against GenBank reference genome 336X in Chapter 5; that the genes associated with carbohydrate metabolism comprise a lower percentage of the overall chromosome. The degradation of D-galactose via the De Ley Doudoroff pathway into pyruvate in *S. proteamaculans* achieves the same goal as the unique itaconate region found in *S. entomophila* isolates. Both pathways, although by differing mechanisms, result in the assimilation of pyruvate. To reiterate from hypotheses proposed in Chapter 4, there is the potential that the replacement of this pathway in *S. entomophila* affords a benefit of sequestering compounds readily available from soil fungi as a form of niche specialisation. To investigate this further, isolate 457 (*S. ficaria*) was also annotated using KofamKOALA, to determine (as a more closely related isolate to *S. entomophila* than *S. proteamaculans*) whether this D-galactonate pathway is complete or also partially lost. *S. ficaria* isolate 457 shows that the De-Ley-Doudoroff pathway is complete in this isolate and also lacks the itaconate degradation cluster from *S. entomophila*, therefore we cannot reject the hypothesis that *S. entomophila* lost this pathway after the acquisition of the itaconate pathway. Mauve alignments of both AGR96X (*S. proteamaculans*) and 626 (*S. entomophila*) further described in section 6.2.5, were used to determine if any further pathways have replaced the De-Ley Doudoroff pathway in *S. entomophila*. The *dgoDAK* gene region in AGR96X is an eight-gene 8,772 bp island flanked 5’ by a hypothetical protein and 3’ *tcdB* dehydratase. These flanking genes are present in 626, although the region between these is approximately 1 Kb in size encoding for a single gene (*phoC*) involved in acid phosphatase activity.

S. proteamaculans isolates encode two additional vitamin and cofactor metabolism degradation pathways (menaquinone biosynthesis and nicotinate degradation) as well as an aromatic compound degradation pathway (homoprotocatechuate degradation). Both the non-pathogenic Chatham Island isolates 440 and MC2 were shown to be the only *S. entomophila* isolate encoding this nicotinate degradation pathway. The *nic* cluster of genes (*nicABCDX*) responsible for the degradation of nicotinate is widely found in soil bacteria, due to plants being the principal source of nicotinate in the environment (Jiménez et al. 2008). Menaquinone biosynthesis *men* gene cluster was identified in all *S. proteamaculans* and found to be lacking one gene block (*menF*) in *S. entomophila.* MenF*,* an isochorismate synthase, catalyses the conversion of chorismite to isochorismate in step 1 of the synthesising sub pathway of 1,4-dihydroxy-2-naphthoate of the quinone metabolism module. Pyrrolnitrin biosynthesis from L-tryptophan was also identified in some isolates *S.. proteamaculans*. Pyrrolnitrin has strong antifungal properties and could be beneficial for a soil-dwelling microbe to synthesise as a competitive advantage in the environment.

S. entomophila isolates were predicted to encode for an additional ATP synthesis pathway module for succinate dehydrogenase when compared with isolates of *S. proteamaculans*. Through manual genome searches of protein sequences, this pathway was later established to be present in *S. proteamaculans* isolates. Genome reduction is often seen by the loss of whole amino acid metabolism pathways. *S. proteamaculans* show the complete module for an extra aromatic amino acid compared to *S. entomophila* isolates. Some isolates also show an additional pathway for cysteine and methionine metabolism, although this is seen across both species. Variation between common regional isolates is also seen. Xenobiotic biodegradation varies between two and three full modules in *S. entomophila*, with no distribution based on geography or pathotype (two encoded in pathogenic isolate 626, and non-pathogenic West Coast vs Canterbury isolates MC2 and 477). *S. entomophila* also uniformly encodes a polyketide sugar biosynthesis module that is only present in some of the assessed *S. proteamaculans* isolates (including 3041 and CfB, of the subspecies *quinovora*). Presence of a complete module isn't uniquely attributed to a subspecies of *S. proteamaculans* however, as AGR96X lacks this predicted pathway module.

For AGR96X especially, modules were explored to find any potential additional metabolic pathways that were not observed in any other isolate, potentially contributing to an increased metabolic capacity of heightened pathogenicity. No identifiable unique KEGG module was found that would give AGR96X a specific metabolic advantage over any other isolate of *S. proteamaculans*. Based on this

finding it was concluded that metabolic pathways for AGR96X are indistinguishable from any other *S. proteamaculans* isolate.

Pathways of the extended *Serratia* species alongside *Y. frederiksenii* Y49 were also predicted to determine anything unique to the *Serratia* family or species *S. entomophila* and *S. proteamaculans* specifically as entomopathogens. Overall, similarities are seen across the *Serratia* genus. All isolates of *Serratia* spp. showed two pathway modules relating to terpenoid biosynthesis, whilst *Y. frederiksenii* showed no complete modules. Of the *Serratia* spp., *S. proteamaculans* show additional modules for vitamin metabolism (14) when compared to *S. entomophila* (12-13). Neither *S. ficaria* nor *S. marcescens* shows the nicotinate degradation or menaquinone complete module, though Y49 has the complete menaquinone biosynthesis cluster. Y49 has, in addition to any *Serratia* isolate, complete modules for pimeloyl-ACP biosynthesis and cobalamin biosynthesis, though lacks a histidine degradation module. For carbohydrate metabolism, other *Serratia* shows a similar number of encoded modules previously described for *S. entomophila* and *S. proteamaculans*, with *S. ficaria* encoding 7 and *S. marcescens* encoding 6 complete modules.

Table 26 Number of complete modules identified by KofamKOALA for metabolic units located on chromosomes of selected *Serratia proteamaculans* and *Serratia entomophila*

* Indicates where results were changed after manual individual genome searches of pathway clusters. Full list of all isolates and their complete pathway modules available in appendix D.2.

Pathway module		*S. proteamaculans*						*S. entomophila*						*S. ficaria*	*S. marcescens*	*Y. frederiksenii*
		AGR96X	**3041**	**143**	**1071**	**10novel**	**CfB**	**626**	**477**	**iDIA**	**A1M02**	**MC2**	**440**	**457**	**1577**	**49**
Carbohydrate metabolism	Central carbohydrate	11	11	11	11	11	11	11	11	11	11	11	11	11	11	9
	Other carbohydrate	7	7	7	7	7	7	6	6	6	6	6	6	7	6	8
Energy metabolism	Carbon fixation	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
	Nitrogen	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
	Sulphur	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
	ATP synthesis	6	5*	5*	5*	5*	5*	6	6	6	6	5	6	5*	5	5
Lipid Metabolism	Fatty acid	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4
	Lipid	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Nucleotide metabolism	Purine	3	3	3	3	3	3	3	3	3	3	3	3	3	3	2
	Pyrimidine	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
Amino acid metabolism	Serine/ threonine	3	3	3	3	3	3	3	3	3	3	3	3	3	3	2
	Cysteine/ methionine	5	5	4	5	4	4	4	4	5	5	4	4	4	4	3
	Branched chain AA	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3
	Lysine	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
	Arginine/ proline	4	4	4	4	4	4	4	4	4	4	4	4	4	5	4
	Polyamine	1	2	2	2	2	2	2	2	2	2	2	2	1	1	0
	Histidine	2	2	2	2	2	2	2	2	2	2	2	2	2	2	1
	Aromatic amino acid	5	5	5	5	5	5	4	4	4	4	4	4	4	5	4
	Other amino acid	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Glycan metabolism	Lipopolysaccharide	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3
Metabolism of cofactors and vitamins	Cofactor and vitamin	14	14	14	14	14	14	12	12	12	12	13	13	12	12	15
Biosynthesis of terpenoids and polyketides	Terpenoid backbone biosynthesis	2	2	2	2	2	2	2	2	2	2	2	2	2	2	0
	Polyketide sugar unit biosynthesis	0	1	0	0	0	1	1	1	1	1	1	1	1	1	0
Biosynthesis of other secondary metabolites	Biosynthesis of other bacterial compounds	2	1	2	2	2	2	1	1	1	1	1	1	1	0	0
Xenobiotics biodegradation	Aromatics degradation	1	3	1	1	1	1	2	2	3	3	2	3	3	1	0

6.2.5 Chromosomal comparisons

For species-wide *Serratia* comparisons orthologous clusters were mapped in a Venn diagram representatives of *S. proteamaculans* (both known subspecies), *S. entomophila*, *S. marcescens*, *S. liquefaciens,* and *S. ficaria* (Figure 57). Between all five species, 3363 orthologous clusters were found. The highest similarity was seen between the two subspecies of *S. proteamaculans,* as expected, closely followed by *S. entomophila* and *S. ficaria*.

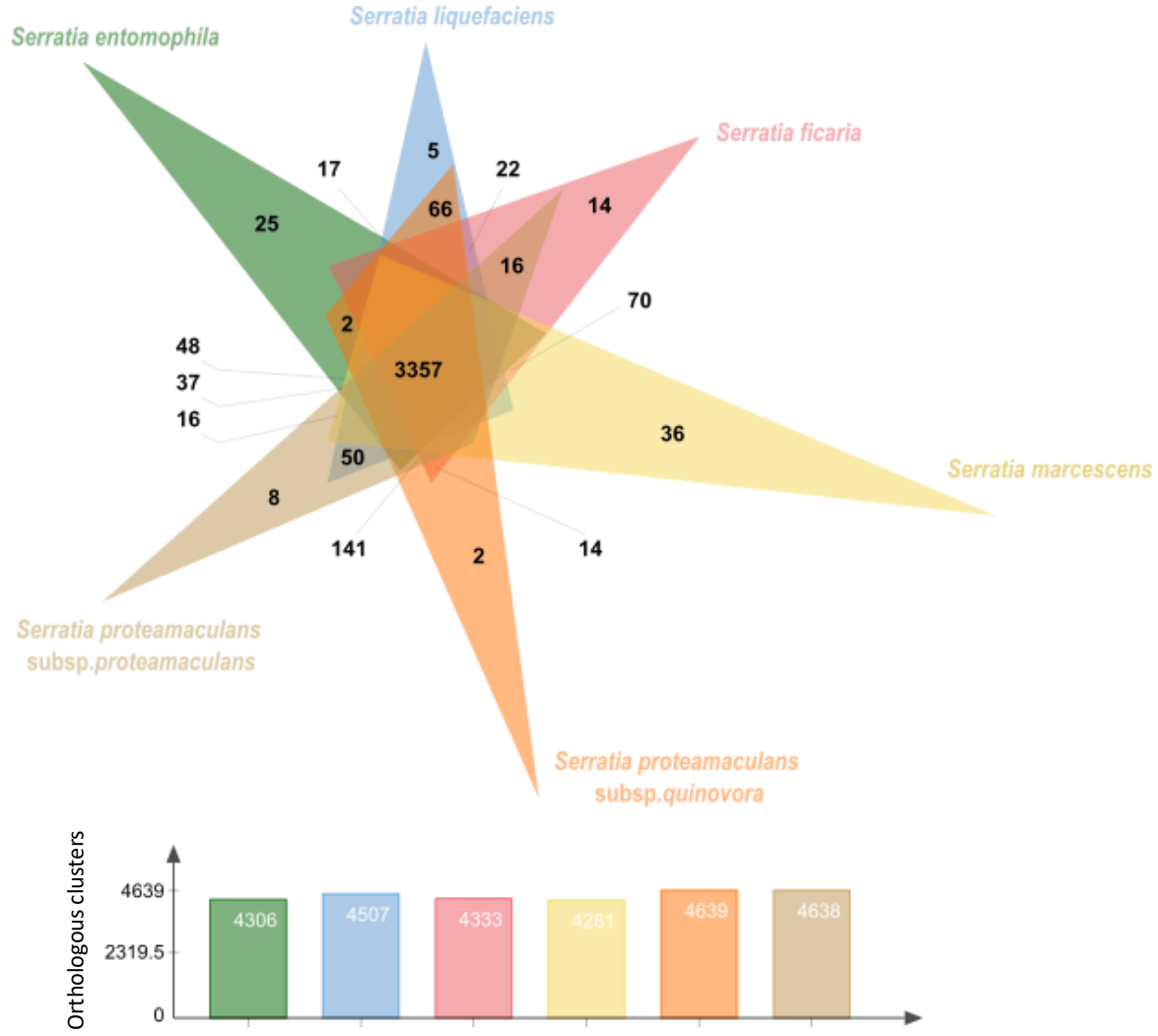


Figure 57 Shared protein clusters in representative isolates of each species of *Serratia* spp. sequenced.
For this representation, AGR96X was used for *S. proteamaculans* subsp. *quinovora,* LC *S. proteamaculans* subsp. *proteamaculans*, 626 *S. entomophila*, 1577 *S. marcescens*, 376 *S. liquefaciens*, and 457 for *S. ficaria.*

Primary genomic comparisons were undertaken on representatives of the two groups of *S. proteamaculans* (subsp. *quinovora* and *proteamaculans*).

Mauve alignments were constructed to determine the differences in collinear blocks between isolates of *S. proteamaculans* and *S. entomophila*. Initial overview alignments were conducted on *S. entomophila* 626, 345, and *S. proteamaculans* 1048, 149, 145, and 143. The general core of the chromosomal layout of these isolates is mostly uniform in the collinear division, although overall the chromosomes of *S. proteamaculans* are larger than *S. entomophila* (Figure 58). Only minor rearrangements and omissions can be observed. Both 1048 and 142 have an extra collinear block, indicated in yellow, which is not shared amongst the other isolates and is inverted on the chromosome of 1048, implying a mobile genetic element. Sequence similarity between these two regions is variable. Both regions share a putative phage sheath protein and are flanked 5' with a *xerD* gene, but 1048 encodes an additional bacillobactin transporter protein amongst the hypothetical predictions. Islandviewer4 identified this region of 626 as a potential genomic island, encoding a fimbrial unit, with partial packaging proteins also found in isolate 345. Isolate 143 shares the same 5' prophage integrase but lacks any of the constructive genes, whilst 1048 and 143 lack even the *intS* integrase gene.

Small blocks comprising of several genes are homogeneous between all three isolates of *S. proteamaculans*, with variable presence in *S. entomophila* (Figure 58). Upstream of the red block is a region encoding a shared tRNA, and two gene green collinear blocks comprising of two hypothetical proteins. The tRNA is also present in *S. entomophila* isolate 345, but the adjacent genes instead encode for three hypothetical proteins in addition to an extracellular serine protease adjacent to a metalloprotease 3' of the variable region (Figure 59). This conformity is similar in 626, though lacking the extracellular serine protease, again showing that variability is even across otherwise similarly conserved isolates.

Another point of differentiation between these chromosomes is between the green and purple collinear blocks, where variable unshared regions are located (Figure 58). The *S. proteamaculans* isolates 1048 and 143 show no additional material at this point, but Mauve alignments (Figure 58) suggest that 342, 145, and 626 encode unique genetic material in this junction point. This region in *S. entomophila* 626 has previously been categorised as a genomic island encoding an extracellular serine protease (section 5.2.9) while this region in isolate 345 encodes additional fimbrial type proteins. Interestingly, 142 encodes an ABC transport complex in this same region that is implicated in phosphate importation, flanked by tRNAs.

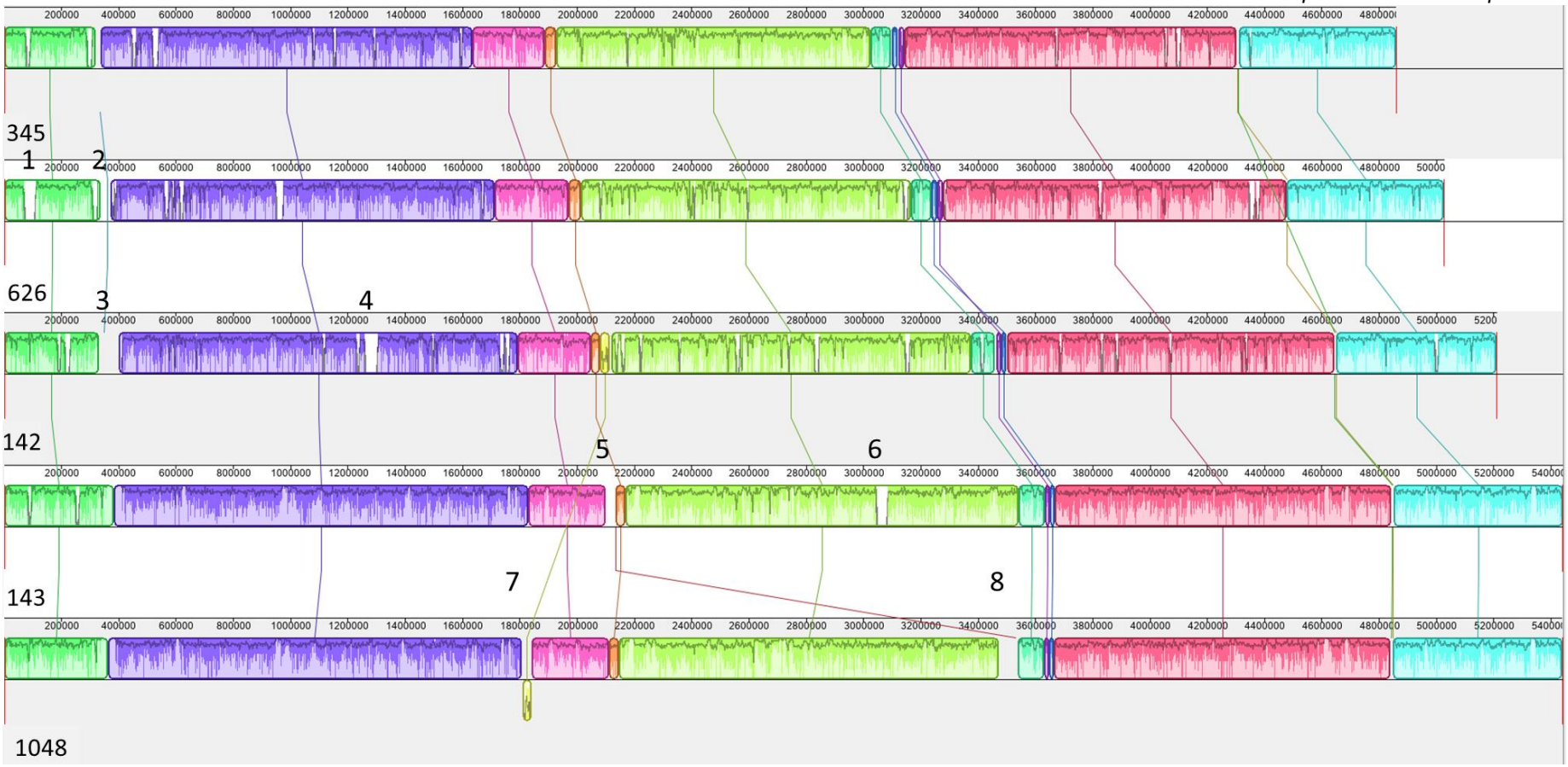


Figure 58 Mauve whole chromosome alignments of two isolates of *Serratia entomophila* (626 and 345) alongside four isolates of *Serratia proteamaculans* (1048, 142, 143)

Similarity skew is shown in collinear blocks. Lines connect similar regions between isolates. Red flanking lines mark the chromosome boundary

1) T1SS amongst hypothetical proteins
2 Extracellular serine protease
3) Phosphate import proteins, transposases, capsule anchoring protein, and chaperone proteins
4) Region of hypothetical proteins
5) Dam methylase protein, XerD proteins
6) *intA* prophage integrase and hypothetical proteins
7) Phage sheath protein and bacillobactin
8) T4SS, cysteine protease, transposases, hypothetical proteins

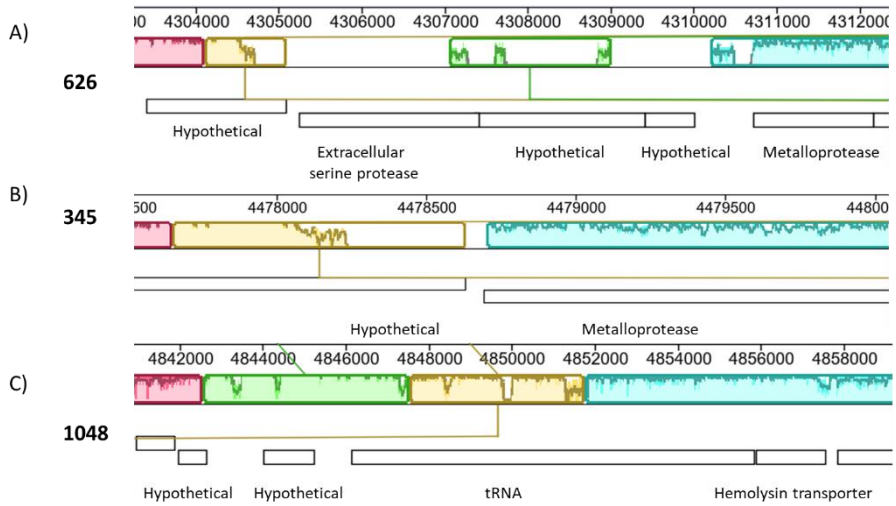


Figure 59 Variable region encoding metalloprotease on *Serratia entomophila*
Genomic Island predicted on all genomes using island finder encoding for different genes. A) and B) *S. entomophila* isolates encoding for extracellular protease and a metalloprotease whereas C) shows *S. proteamaculans* 1048 only having the tRNA (ubiquitous across isolates of *S. proteamaculans*).

Overall, when comparing the conformity of the genome between these two species, there are high levels of homogeneity. Large regions are conserved, with minor additions or reductions in conserved regions between the two species. In relation to *S. entomophila* and *S. proteamaculans,* the latter tend to have more unique regions. Small unique regions are seen in isolate 626 based on BRIG's alignments displayed earlier (Figure 54) and this was to be expected. Rearrangements in the *S. proteamaculans* isolates seem to be more common, with inversion of the yellow collinear block as a prime example (Figure 58).

Using *S. proteamaculans* AGR96X and *S. entomophila* 626 as model genomes for their species, whole-genome alignments were undertaken, and regions of difference were compared and annotated for each chromosome. Overall, the chromosomal configuration is mostly uniform for both isolates. A small region, identifiable in Figure 60 (dark green), in both 626 (at approximate position 4.4 Mb on the genome) and AGR96X (position ~0.55 Mb) is translocated between two different orthologous groups. Another rearrangement observed is where a conserved block (purple) shows some points of interest between these two isolates that are identifiable chromosomal differences. This conserved block appears on the 'reverse' strand of *S. entomophila* as opposed to the primary strand for *S. proteamaculans* AGR96X, indicating a chromosomal rearrangement. This block encodes for an enterobactin siderophore module. From the analysis of predicted genomic islands in *S. entomophila*

626, this enterobactin module is 5' of a large genomic island comprising of a 50 Kb region (marked by region 5 in Figure 60). Insertion of this island could have caused rearrangements to the existing chromosome that allowed for the inversion of the enterobactin region. A T1SS was observed on the chromosome of 626 at position 4.2 Mb. BLASTn results show this operon has 77% identity to an *S. marcescens* ABC exporter system gene cluster. This region again was identified previously in this study in isolate *S. entomophila* 626 as a predicted genomic island (Chapter 5, section 5.2.9).

In comparing alignments of representative *S. proteamaculans* isolates (of both subspecies) greater variability in the chromosome is observed than seen in *S. entomophila*. From these observations on the chromosome alignments when compared to predicted islands, one of the largest causes of genomic rearrangements specifically in *S. proteamaculans* or areas of heterogeneity is the acquisition, or lack thereof, of horizontally acquired material.

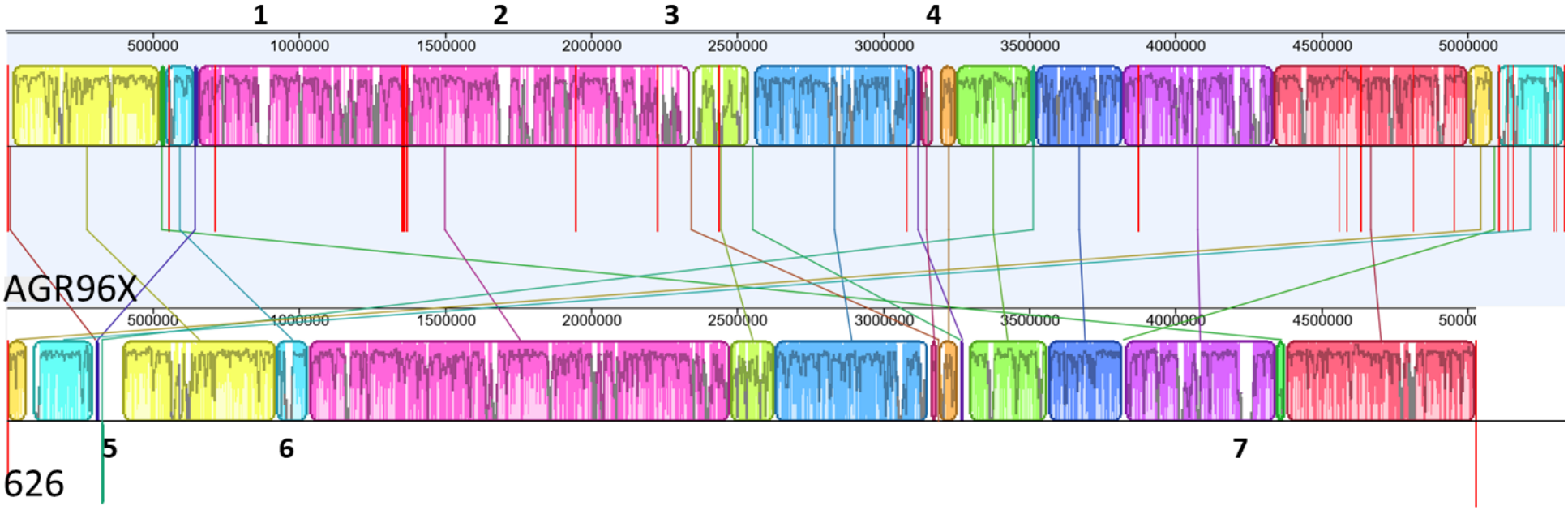


Figure 60 Overall comparison of the chromosomes of *Serratia proteamaculans* AGR96X (top) and *Serratia entomophila* 626 (bottom) highlighting points of interest
1) Ribose import gene cluster rbsABC alongside a carboxylesterase and hypothetical proteins
2) Putative prophage major tail protein, hypotheticals, lysozymes xerD
3) Hydrogenase maturation cluster, hydrogenase 3 maturation protease, nickel transport unit
4) Multiple transcriptional regulators, aligns with predicted genomic island in *S. entomophila*
5) Large putative genomic island encoding an extracellular serine protease, multiple transcriptional regulator proteins and virulence sensor protein BvgS. Green line below the central
chromosomal line indicates the rearrangement of the enterobactin encoded genes.
6) Heme binding protein, hypotheticals and outer membrane assembly protein
7) Polyketide synthase, T1SS

As previously discussed in Chapter 5, the ability of *S. entomophila* to degrade itaconate is used in a selective medium to differentiate between *S. entomophila* and *S. proteamaculans*, which cannot grow on itaconate agar. Chromosomes of all sequenced *S. proteamaculans* were searched for the previously discussed *S. entomophila* itaconate degradation cluster identified comprising of a transcriptional regulator with 3 functional genes (*mcl, mch, yfdE*) and a transporter protein, to establish whether any partial island acquisition had occurred. Neither the operon nor transcriptional regulator protein was found in any of the *S. proteamaculans* isolates (Figure 61). At the 5', flanking *pip* gene and 3' *holE* was identified in genome alignments in both isolates of *S. entomophila* and *S. proteamaculans*, but the itaconate genomic island itself was shown to have not been acquired. In all *S. proteamaculans* isolates, a 29 bp intergenic region is found between these two genes where the itaconate degradation cluster is found in *S. entomophila* isolates. No tRNAs were identified flanking the *S. entomophila* itaconate region. The itaconate degradation cluster was then searched for in *S. ficaria* isolate 457, as this is the most similar species to *S. entomophila* as deduced from ANI and phylogenetic analysis. The region in 457 matches what is found in isolates of *S. proteamaculans*, implying the acquisition event that allowed *S. entomophila* to gain the itaconate degradation cluster occurred after the common ancestors that split from *S. proteamaculans* and subsequently *S. ficaria* (region further assessed in Chapter 5, section 5.2.11).

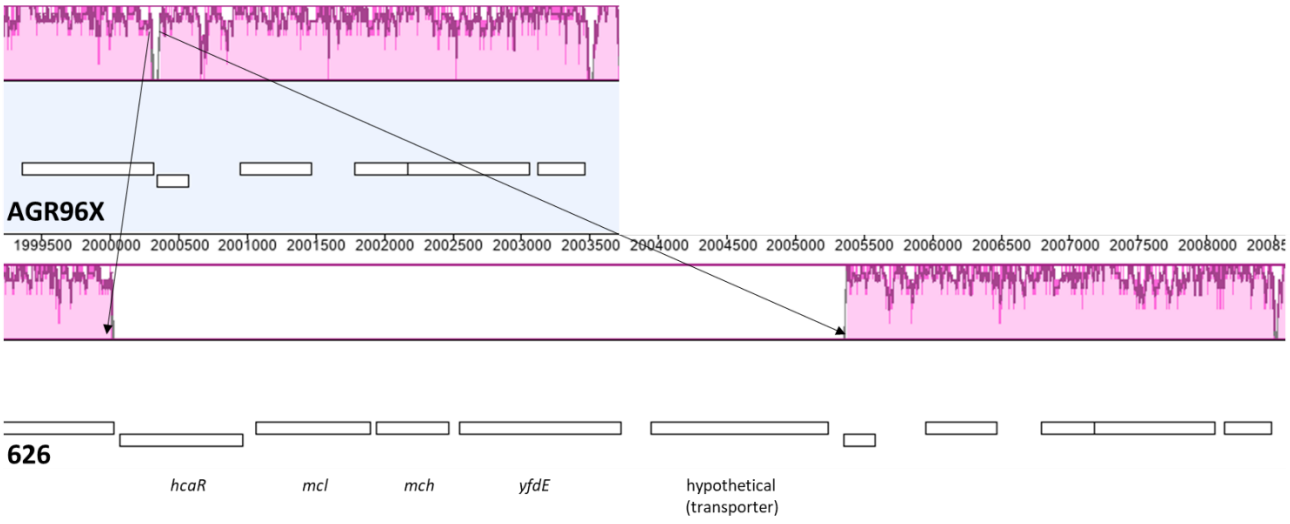


Figure 61 Itaconate region of *Serratia entomophila* missing from *Serratia proteamaculans* AGR96X
Itaconate degradation operon encoding *hcaR* transcriptional regulator, *mcl, mch* and *yfdE* alongside a transporter protein predicted with Phyre. The region is flanked by 5' pip and 3' holE uniformly found in *S. entomophila*, *S. proteamaculans,* and *S. ficaria*.

6.2.6 Detection of genomic Islands and regions of gene acquisition

6.2.6.1 Prediction of putative genomic islands in *S. proteamaculans* isolates

To further investigate the proclivity of *S. proteamaculans* for horizontal gene transfer and validate the differences found using Mauve, prediction of potential genomic islands was undertaken on isolates of hyper-, chronic, and non-pathogenic *S proteamaculans* to ascertain any differences and compare them to that of *S. entomophila* (Figure 62).

Hypervirulent isolate AGR96X has a greater prevalence of potential islands on its chromosome relative to 3041 and 145. Although genomic islands are observed to be more abundant in 3041 than isolate 145, the chromosomal position of these genomic islands is more similar between these two isolates than with AGR96X. The G+C skew of these three isolates matches the predicted genomic islands. AGR96X has several prominent spikes (Figure 62), although a region located close to the 5 Mb region on the chromosome has the longest span of elevated G+C. A list of predicted genomic islands for AGR96X are listed in Table 27.

In comparison to the predictions of *S. entomophila* isolates 626 and 477 discussed in Chapter 5, section 5.2.9, many differences are observed between *S. entomophila* and *S. proteamaculans*. *S. entomophila* 626 has increased predicted genomic islands and is similar in numbers to that of *S. proteamaculans* 145.

Table 27 IslandViewer4 hits predicted *in Serratia proteamaculans* hypervirulent isolate AGR96X
Genes listed are those predicted to have a function or are virulence/ TA systems that may or may not have a regulatory function when inserted into the chromosome.

Chromosomal position			
5'	3'	Size (bp)	Genes located on a predicted island
1	5895	5894	Cation acetate symporter *actP*
12976	21377	8401	Pyrrolnitrin synthase
850385	882150	31765	Prophage integrase/ hypothetical
1001717	1007287	5570	Fimbrial unit
1143358	1147414	4056	Pyramidine ribonuclease
1345015	1362643	17628	Mycothiol biosynthesis pathway associated genes
1376015	1383875	7860	*tagH* energy coupling gene
1699286	1721087	21801	Hypotheticals, Dam methylase,x*erD*
1895100	1900007	4907	Fructose metabolism
2051090	2057383	6293	Hypotheticals flanked by tRNAs
2223797	2236022	12225	Hypotheticals
2262869	2267619	4750	Transport system
2348298	2373692	25394	Fimbrial unit/ vitamin B12 transporter
2414615	2426127	11512	Beta-ketoadipate pathway (aromatic compound)
2544064	2557441	13377	Iron uptake
2571518	2577140	5622	Phage shock complex ABC
2645966	2651785	5819	Hypothetical
2836584	2852741	16157	Transposase, fimbrial proteins
2842212	2851346	9134	Nitrate assimilation cluster
2990384	3005103	14719	Hypotheticals, protein phosphatase
3005120	3019339	14219	Hypotheticals, partial TA systems VgrG1/ClpV1
3112061	3116885	4824	Hypotheticals
3148721	3153519	4798	Fimbrial and chaperone proteins
3169408	3190667	21259	*kefF*, transcriptional regulators
3694229	3701275	7046	Transcriptional regulator, sensor kistidine kinase
3811171	3842919	31748	Prophage integrase, *pasI* antitoxin, inner membrane proteins, hypotheticals
4194322	4207255	12933	Fimbrial proteins, hypotheticals
4265129	4269639	4510	Type II secretion system
4429226	4433392	4166	tRNA nuclease, hypotheticals
4732476	4740498	8022	Hypotheticals, transcriptional regulator, putative protein
4903043	4909374	6331	Transporter proteins
5084477	5117427	32950	Prophage integrase, transcriptional regulator, fimbrial proteins, hypotheticals
5358541	5363390	4849	Transcriptional regulator, sensor kistidine kinase
5367787	5373245	5458	Transposase
5377044	5381854	4810	Lyzozyme inhibitor, transcriptional regulator
5390866	5398465	7599	biofilm synthesis proteien *pgaD*
5405357	5413071	7714	Fimbrin specific adhesin *fimH*, transport nit
5456248	5474113	17865	Transcriptional regulator, hypothetical

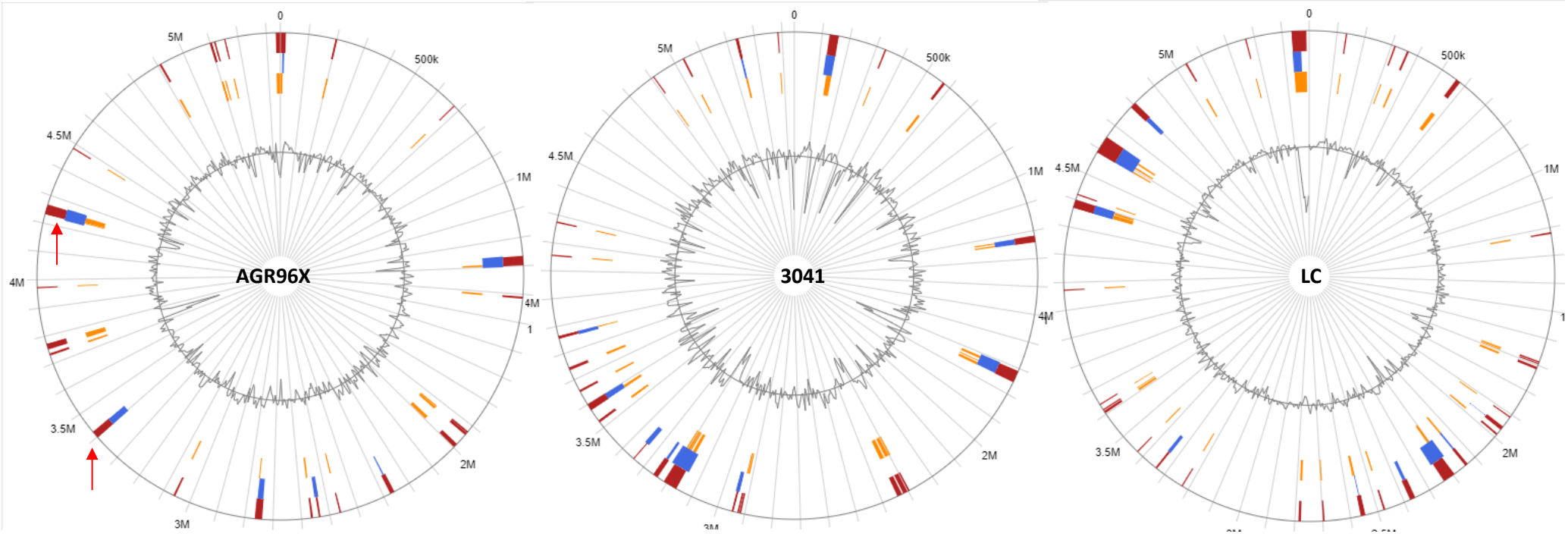


Figure 62 Predicted genomic islands on the *Serratia proteamaculans* chromosome *of Serratia proteamaculans* subsp*. quinovora* A) hypervirulent AfpX encoding AGR96X B)Non-pathogenic 3041 and C) pathogenic *Serratia proteamaculans* subsp. *proteamaculans* LC
Red indicates that the island was predicted by at least one method, where blue (Islandpath-DMOB) green (IslandPick) and orange (SIGI-HMM) indicated individual prediction methods. G+C skew shown in the centre circle. Red arrows indicate putative phage's on AGR96X.

6.2.6.2 Prediction of chromosomally encoded transposases

In addition to the prediction of genomic islands, ISFinder was used to identify transposases on the chromosomes of *S. proteamaculans* isolates to identify potential sites of horizontally acquired inserts (Table 28). The number of predicted horizontally acquired transposases is much larger than in *S. entomophila*. Both species share ISkRa4 (ISCep1) and Tn3 (TnXaX1) as common transposases across isolates in a mostly conserved region, implying these were acquired from a common ancestor of both species. Through chromosomal searches of *S. liquefaciens* and *S. marcescens* isolates, these two previously mentioned transposases were also found. Tn3 (TnXaX1) was also found in genome searches of *Y. frederiksenii* isolate Y49.

In *S. entomophila* horizontal acquisition events could be predicted more accurately due to the low number of dissimilarities in transposase orientation. This was not observed in *S. proteamaculans*, where multiple isolates would share a similar genome inversion, meaning the origin of the acquired event could not be predicted. Instead, HGT events were recorded by how many different orientations per transposase were observed. Overall, 10 transposases showed multiple chromosomal orientations. Of these, several had more than one alternate location, including ISkRa4 ISCep1 (Table 28). From this, the gene acquisition events for the translocation of these transposases, and therefore HGT events overall, occur more frequently in *S. proteamaculans* than in *S. entomophila*. The frequency of these events implies that *S. proteamaculans* has a greater opportunity (or susceptibility) than *S. entomophila* to acquire new transposable elements as a source of potential adaptive evolution.

Table 28 Predicted transposases found on the chromosome of *Serratia proteamaculans* isolates using ISfinder

6.2.6.3 Prediction of restriction-modification and CRISPR-Cas systems

In addition to the prediction of regions acquired through horizontal means, detection analysis for systems utilised by bacteria to evade these events was undertaken. Restriction modification (RM) systems and CRISPR-Cas arrays are known from the literature to be inhibition and protective systems against bacteriophages and conjugation/ transformation events respectively. The detection of these systems (or lack thereof) in isolates of *Serratia* would indicate any preventative methods bacterial isolates have to resist the acquisition of foreign DNA.

CRISPRcasFinder web tool was utilised to search chromosomes of all *Serratia* genomes to identify the presence of any CRISPR-Cas systems. These systems have the potential conferred immunity to plasmid or foreign DNA based on the RNA spacer sequence. Due to this, the presence of CRISPR-Cas arrays has been implicated in bacterial adaptive immunity, potentially reducing the transfer of mobile genomic elements and bacteriophage invasion. Predictions of CRISPR-Cas arrays found that CAS I-E were found in less than half the isolates of *S. entomophila* (Table 29). These were predominantly found in non-pathogenic isolates, including West Coast and Chatham Island isolates (*S. entomophila* MC2, 440, 442, and 294). CAS I-E is a type I system encoding a single operon of single copy CAS-associated genes. This CAS system has been previously described in *E. coli* K12 (*ygcG- ygbF*), commonly found in members of Enterobacteriaceae (Xue and Sashital 2019).

Table 29 Prediction of CRISPRcas arrays in *Serratia* isolates
The number of arrays encoded per chromosome indicated in brackets.

CRISPR-Cas system	*Serratia*	Genome detected
CAS I-E	*entomophila*	440(3), 442(3), 294, MC2(3), 345(3)*, 219(3)*, 176(3), 158(3), 210(3), 1100(3)
CAS I-F	*proteamaculans*	12newD(2)

Where evidence level was indicated as 'low' by CRISPRcasFinder, arrays were left out of the total presented in Table 29. Since low evidence does not rule out the presence of an active CRISPR array entirely, these arrays are indicated by *.

For RM systems, manual genome searches of MTase (methyltransferase) genes were undertaken. These were then allocated as potential restriction-modification (RM) systems if a REase (restriction endonuclease) was located within four CDS on either side of the MTase. In the case of predicted hypotheticals, BLASTn was undertaken to determine whether this was potentially a REase. If no discernible hit was found, then this could not be ruled out as an RM system and was included in the results.

In *S. entomophila* isolate 626, two type I RM REase genes and one type II REase gene were identified. The type II REase (apIIR) was found to be adjacent to modification methylase paeR7IM. The type I REase genes, *hsdR* and *hsdM* were found in a single operon ordered *hsdM*, hypothetical, *hsdR*. BLASTn of the hypothetical protein reveals a type I restriction enzyme specificity protein, as specified in Uniport, where type I systems are comprised of RMS polypeptides. The assembly of this restriction-modification system aligns with that described by *Helicobacter pylori*, resulting in one complete type I RM system, in addition to one type II system. This *hsdRSM* operon was additionally found by BLASTn in *S. entomophila* isolates 477 and A1M02, with Chatham Island isolates 440, 442, and 294 alongside MC2 showing a translocated operon running on the reverse strand. No isolate *of S. proteamaculans* returned any results for this type I RM system. Of the extended *Serratia* genus isolates, only *S. ficaria* returned a BLASTn hit for this type I RM system.

Initial searches of AGR96X hypervirulent *S. proteamaculans* revealed that a type II RM system could be located on the chromosome, comprising of NotI adjacent to modification methyltransferase haeIIM. No Type I RM systems were found in AGR96X.

6.2.6.4 Presence of putative phages

Functional and non-functional phage-associated genes have been described widely in bacterial genomes and have been implicated in significant variation between species. Acquisition of antibiotic resistance or adaptation to a niche has often been associated with these phage-associated genes, therefore predicting them for comparisons between *S. entomophila* and *S. proteamaculans* genomes in addition to the wider *Serratia* spp. was pertinent to further describe the acquired differences between these two species.

On the assessment of the *S. proteamaculans* chromosomes, there were no phage predictions that were uniform to all isolates (Table 30). The most frequently found phage cluster was an ~22 Kb (intact) region predicted to be most similar to phage PHAGE_Erwini_ENT90_NC_019932. Isolate 3041 encoded the highest number of complete prophage regions (5). LC, 299, and 336 all predicted six (either intact or incomplete) phage-like regions. Phage predictions for *S. entomophila* revealed that overall, the number of intact and incomplete phage predictions were drastically reduced in comparison to that of *S. proteamaculans*, except for isolates sourced from the Chatham Islands, West Coast isolate MC2, and Man3 (Table 30). Reduced acquisition of phage regions is another indicator that the genomic diversity in isolates of *S. entomophila* is lower, except for Chatham Island isolates which, based on predicted genomic islands and presence of transposases, have a potentially higher

rate of gene acquisition events. Uniformity was observed in the prediction of a ~19 Kb phage-like region across all isolates of *S. entomophila* (*in silico* observations explored *in vitro* in Chapter 5, section 5.2.8). BLASTn of this region revealed high similarity to a region found in *S. marcescens* isolates (84.79% identity to 80% cover), with no hits to *S. proteamaculans*. This region encodes DinI damage-inducible protein alongside motifs for two SOS box binding sites, lysozyme *rrrD,* and hypothetical proteins when annotated with Prokka. Annotations of the genome in RAST reveal these hypotheticals to be phage proteins. This phage region is varying in size and completeness across the isolates of *S. entomophila*, though all carry the region flanked by tRNAs and carry *dinI.*

Table 30 Phage predictions of intact and incomplete phage regions in all isolates
Predictions made using Phaster.

		Phage predictions		
Species	Isolate	Intact	Incomplete	Questionable
S. proteamaculans	AGR96X*	2	2	0
	4[1]	2	1	0
	10Novel*	2	1	0
	12a*	3	3	1
	12newD[1]	2	0	0
	25e*	2	1	0
	142[1]	1	1	0
	143[1]	2	1	1
	145*	3	1	0
	149[1]	3	1	1
	299[1]	3	3	0
	336[1]	3	3	0
	465[1]	2	0	1
	1048*	2	0	0
	1071[1]	2	1	0
	1129[1]	1	4	1
	1137*	2	0	1
	1457[1]	2	0	2
	1769*	3	0	0
	3041*	5	0	0
	20093*	2	1	0

	CfB[1]	2	1	0
	G[1]	1	3	0
	K[1]	2	2	0
	LC[1]	4	2	1
	M[1]	2	1	0
	Puna18*	0	2	30
S. entomophila	626	2	1	0
	158	1	0	0
	176	1	0	0
	210	1	0	0
	219	0	0	1
	220	0	1	1
	294	4	1	0
	305	0	0	4
	345	1	0	0
	364	2	0	0
	398	0	0	2
	440	3	0	1
	442	3	1	0
	477	2	1	0
	482	0	0	0
	1100	3	0	0
	1343	0	1	4
	A1M02	0	0	3
	I	0	2	0
	iDIA	0	1	2
	Man3	5	2	0
	MC2	3	0	0
S. liquefaciens	F28	0	8	0
	377	1	2	0
	376	1	2	0
S. marcescens	1577	5	2	0
S. ficaria	457	1	1	0

* Indicates subspecies *quinovora* in *S. proteamaculans*

[1] Indicates subspecies *proteamaculans* in *S. proteamaculans*

6.2.7 Pangenome wide association and core genome comparative analysis

As was previously undertaken with *S. entomophila* (Figure 31), pangenome analysis was undertaken using Roary (Figure 63). Using the 28 sequences of *S. proteamaculans*, core genes were identified in addition to unique genes on the chromosome (unique being defined as genes only encoded on one isolate sharing <95% sequence similarity). Comparing statistics, the pangenome of *S. proteamaculans* is comprised of 3121 conserved genes (Figure 63). These are mostly uniform across all isolates, with 1137 showing the greatest deviation and small levels of unique genes in many other isolates (under 1000 genes). Isolates 336, 25e, and 10Novel are the only isolates that show no unique genes. Isolate 1137 also shows high levels of deviation from other genomes with approximately 1000 unique genes. The presence of unique genes is highly variable in this dataset when compared with *S. entomophila.*

The largest number of unique genes present in any assessed *S. entomophila* isolate was 123 in iDIA (Chapter 5, section 5.2.4). The core genome of *S. entomophila* was approximately 750 genes larger than that of *S. proteamaculans* (for both subspecies *quinovora* and *proteamaculans*), although individual isolates had similar numbers of CDS predicted on the chromosome. This greater level of deviation among *S. proteamaculans* isolates again shows that genetically they are more variable and have acquired various genomic traits from other organisms.

Figure 64 shows the compiled shared pangenome of both *S. entomophila* and *S. proteamaculans*, alongside additional species of *Serratia* and the *Y. frederiksenii* isolate Y49, visually demonstrating the differences in pangenome homogeneity between the two species. *S. entomophila* isolates share a large, conserved block (as also seen in Chapter 5, Figure 31). The addition of 28 *S. proteamaculans* isolates display the shared differences that account for the non-unique but non-pangenome genes present. Five isolates (Puna18, 3041, CfB, 145, AGR96X) also share a block of similar genes in the accessory genomes. On investigation through searches of the Roary gene/presence output file, these similar genes mostly return as hypothetical proteins. In addition to hypothetical proteins, there are regional operons present that are lacking transcriptional regulators. A key gene of interest present in these five isolates is *pgdA*, a gene encoding a polysaccharide deacetylase family protein 89.4% AA similarity to *S. liquefaciens*. Functionally, deacetylase family proteins can be involved in the deacetylation of chitin. This action has been described in *Streptococcus pneumoniae* as contributing to virulence by conferring host lysozyme resistance (Vollmer and Tomasz 2002). This region has no relationship to pathogenicity since all three pathotypes described are represented in this group. Assessing differences based on groupings allocated in Chapter 3, group 1 (*S. entomophila* and *S. ficaria*) cluster separately to *S. proteamaculans* isolates (indicated with a black box in Figure 64). In distinguishing subspecies boundaries, the lines become a bit more blurred. Isolates AGR96X, 145, CfB

Puna18, and 3041 clearly show larger regions of homogeny when compared to other isolates. This is indicated more by the dendrogram to the right of Figure 64, where these five isolates group separately to other *S. proteamaculans* and more closely to *S. liquefaciens*. Other assigned members of the subspecies *quinovora* such as 20093 and 12a do not align so closely as inferred by 16S phylogeny. This corresponds to the ANI values reported in section 6.2.1 (Figure 51) where only the five isolates (AGR96X, 145, CfB Puna18, and 3041) are grouped accordingly to be classified as a similar subspecies.

With reference to Figure 64, there are no identifiable blocks solely attributed to the AfpX encoding isolates (20093, AGR96X, LC, MH5). One conserved region is unique to AGR96X. By manipulating the Roary output .csv file it was found that AGR96X encodes for a chitoporin (ChiP_1) not found in any other *S proteamaculans* isolate. The region with ChiP_1 is 10,531 bp flanked with tRNAs. This region does not differ from the average G+C content of the chromosome. Each chromosome of *S. proteamaculans* encodes for 1 chitoporin gene, but the addition of a second 1 chitoporin gene has implications on the uptake of chitooligosaccharides, and therefore the degradation of chitin (Aunkham et al. 2018)

Scoary analysis was undertaken before determining if any genes could be positively identified as attributed solely to plasmid-bearing isolates. Unfortunately, no gene correlation could be found.

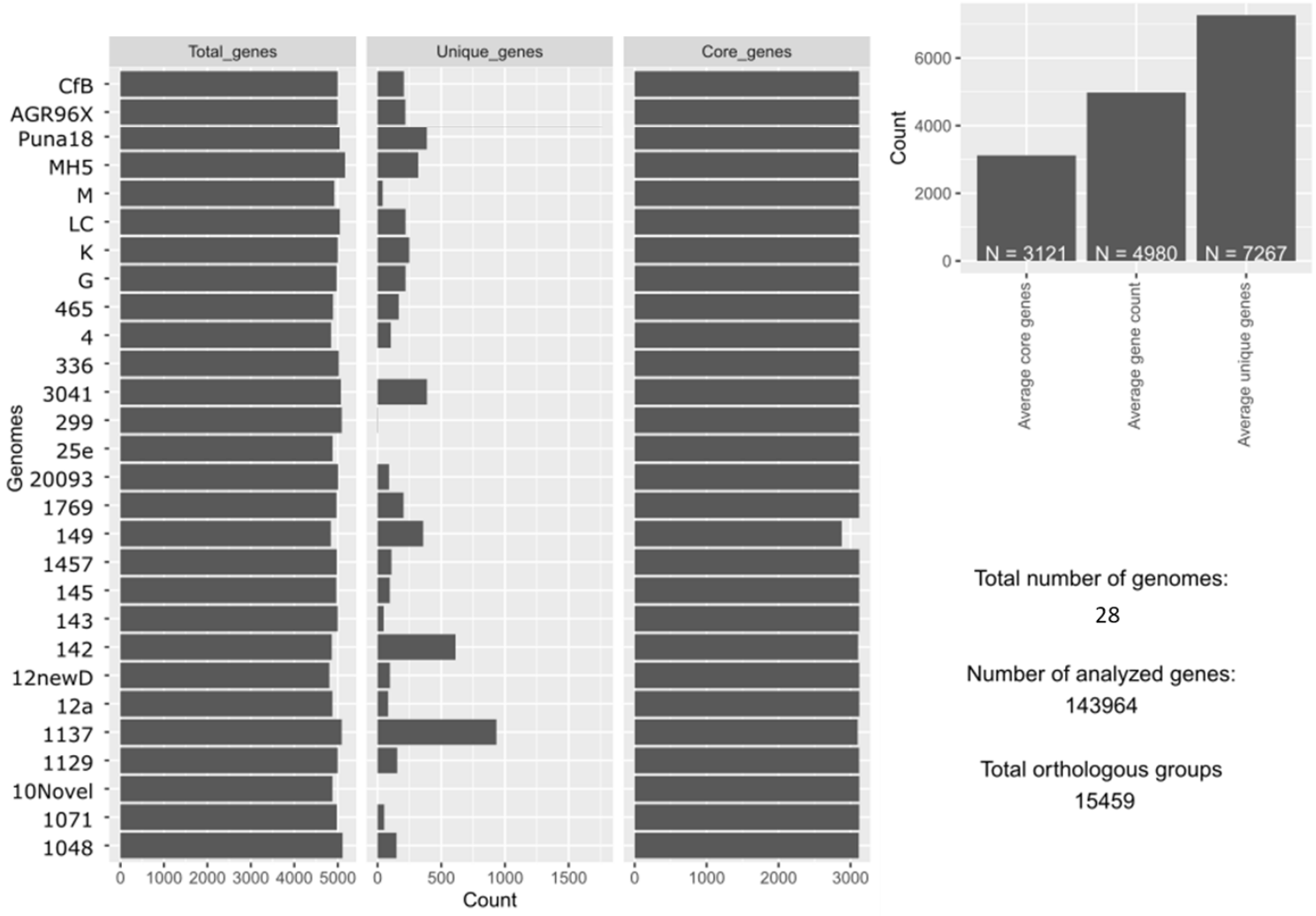


Figure 63 Pangenome of isolates of *Serratia proteamaculans*
Total genes refer to the overall number of predicted genes per isolate. Unique genes are genes only associated with one isolate and core genes are genes shared across all isolates of *S. proteamaculans* included in this study.

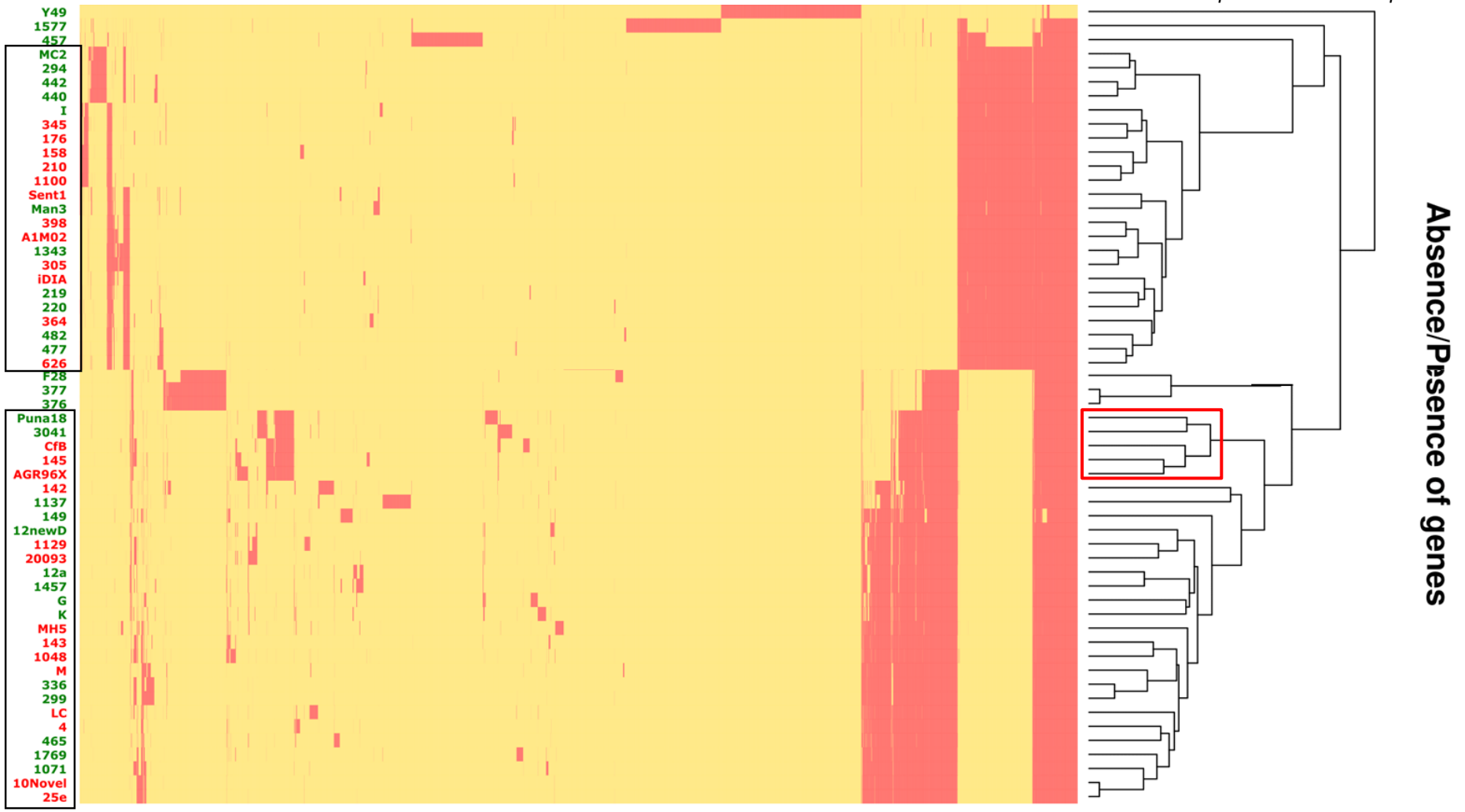


Figure 64 Presence and absence graph of genes in all sequenced isolates
Yellow indicates absence whereas red indicates a shared gene across more than one isolate. Dendrogram based off the core genome. Isolate name in green indicates non-pathogenic isolates whereas in red indicates pathogenic (of any degree). Black boxes indicate species, where the top box composes isolates of *S. entomophila*, and the bottom shows *S. proteamaculans*. Non-boxed isolates are extended *Serratia* spp. and isolate Y49 of *Y. frederiksenii*. Isolates 1577 and 457 are *S. marcescens,* whereas F28, 376 and 377 are isolates of *S. liquefaciens*. Red box outlines potential 'core' delineation of *S. proteamaculans* subspecies *quinovora*.

6.2.8 The openness of the *S. proteamaculans* pangenome

In Chapter 5, the openness of the *S. entomophila* pangenome was discussed. Orthologous gene clusters were calculated for all 28 isolates of *S. proteamaculans* before the pangenome openness was calculated using Micropan. Here, the pangenome of *S. proteamaculans* is analysed following the same method as in Chapter 5. As discussed earlier, an open genome can be characterised by the perpetual increase of unique genes with the subsequent addition of extra genomes to the assessment criteria. Rarefaction curves of *S. proteamaculans* pangenome show a heavy increment in unique new genes to genomes added (Figure 65A), in contrast to what was observed in *S. entomophila* (Figure 65B).

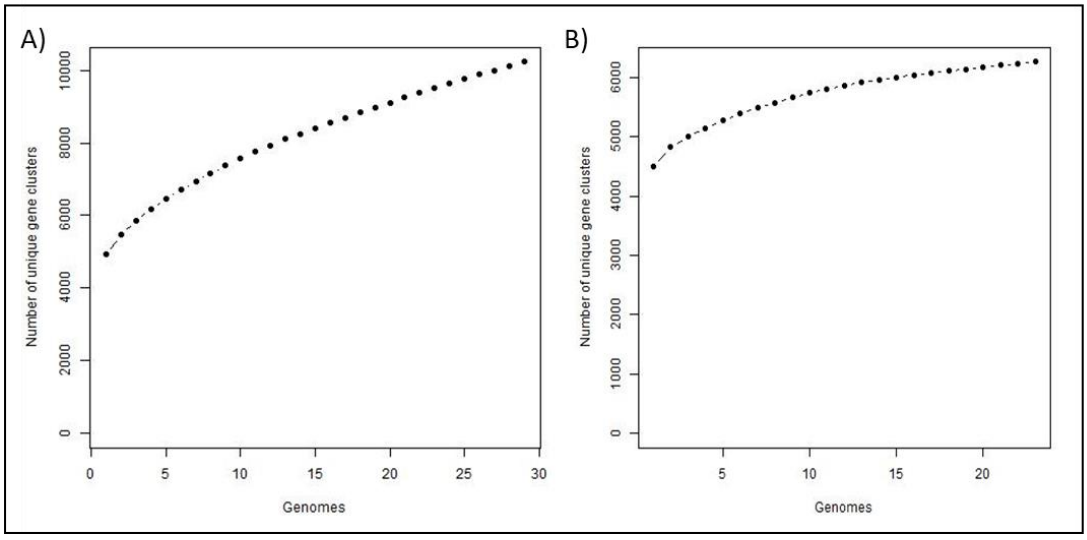


Figure 65 Rarefaction curve of *Serratia proteamaculans* pangenome gene cluster analysis compared with *Serratia entomophila* made in Micropan
Panel A) shows *S. proteamaculans*, whereas panel B) shows *S. entomophila.*

With 1000 permutations, the alpha value of Heap law was calculated to be 0.566, meaning the population level is unbounded. This number, however, is significantly lower than that of *S. entomophila* (0.94). Although *S. entomophila* pangenome is open, the level of new unique genes in *S. proteamaculans* is significantly larger, potentially showing the large range of acquired genes and clusters on its chromosome. Utilising 28 isolates to determine this alpha number should provide an accurate insight into the openness of the genome. This validates the findings of increased levels of genome variation and potential for horizontally acquired regions in isolates of *S. proteamaculans* shown in section 6.2.6. Further analysis using Chao's estimate calculation set the lowest bound of the complete pangenome size to 17684 gene clusters. This value is 2.6x the estimated lower bounds of *S. entomophila* pangenome. These results show that although the pangenome of both species is open, *S. proteamaculans* has more opportunity and history of gene acquisition from the environment. This

supports the hypothesis that *S. proteamaculans* has high levels of diversity on the chromosome, potentially from the higher rates of horizontally acquired DNA.

To again confirm the subspecies groupings previously classified, pangenome analysis was undertaken on a cohort of isolates after the removal of known *S. proteamaculans* subsp. *quinovora* isolates, leaving *S. proteamaculans* subsp. *proteamaculans* (Figure 66). The resulting Heap value returned 0.557 and a Chao estimate of 14,346 genes, showing little difference when compared to the whole cohort. This demonstrates that, overall, the pangenome does not differ greatly in size between the two subspecies, though extra genes are noted in the combined total.

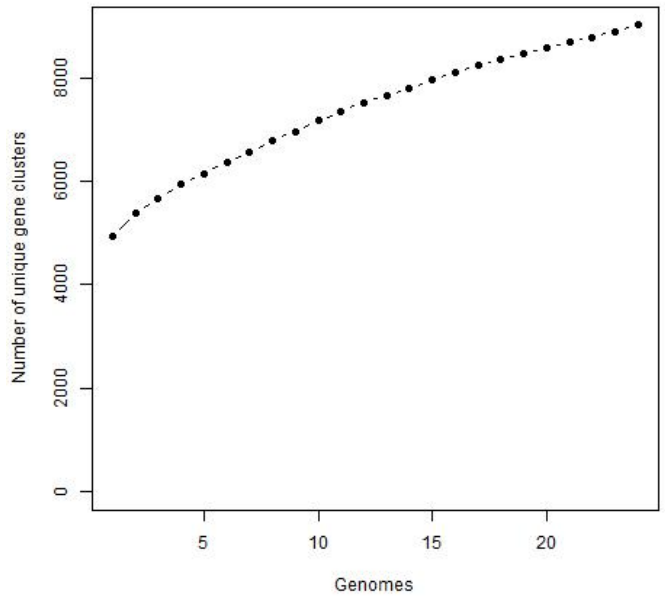


Figure 66 Rarefaction curve *of Serratia proteamaculans* subsp. *proteamaculans* pangenome gene cluster analysis made in Micropan

6.2.9 Chromosomally encoded accessory virulence determinants

As conducted in Chapter 5 linking to the results of plate assays found *in vivo* in Chapter 4 and transconjugant bioassays in Chapter 7, identification of chromosomally encoded accessory virulence determinants was undertaken. For this analysis, it was not expected there would be any difference between subspecies, given the similarities observed between them in Chapter 4. Specifically looking at the detection of proteolytic enzymes (peptidases) in hypervirulent *S. proteamaculans* AGR96X would reveal differences not previously seen. As previously detected in *S. entomophila*, Chitinase A, B, and A1 alongside a hypothetical protein attributed as a lipoprotein were detected. No motif hits were found with Pfam *glyco_hydro_family_19*.

Chitin binding protein search was utilised with the Pfam family LMPO_10. AGR96X showed the presence of two chitin-binding proteins (GlcNAc binding protein A) (Table 31). These two binding proteins were identified as being synonymous with the *S. proteamaculans* binding proteins identified on the CAZY database with 95.43% (accession ABV42576.1) and 86.2% (accession ABV43333.1) amino acid (AA) identity respectively. One of these chitin-binding domain proteins also shares 93.37% AA similarity with the domain found in *S. entomophila* isolates. Searches of the wider cohort of *S. proteamaculans* isolates found various degrees of encoded gene copies of the GbpA binding protein across isolates, ranging from 1-2 copies. Six isolates were encoded for one copy of the *gbpA* gene (4, 142, 1457, 3041, 12a, CfB), whilst the rest of the *S. proteamaculans* isolates encoded two copies. No correlation was found between copies encoded or virulence. General HMM searches of the wider *S. proteamaculans* cohort showed that chitinase A, B, and A1 present across all isolates. Searches using the enzyme entries for EC 3.2.1.14 identified four chitinases on the chromosome of AGR96X again as well as two chitin-binding proteins.

Further analysis of the chromosomes of other isolates for these chitinases and chitin-binding proteins we see increased variability. Chitin binding protein copies are encoded for up to three copies on a genome, with the majority of isolates encoding three copies.

Table 31 HMMER results for chitinase glyco-hydro family motif search and chitin-binding protein family on *Serratia proteamaculans* isolate AGR96X with percentage similarity of protein sequence to the most similar protein amino acid sequence on UniProt.

Protein name	Function	Size (AA)	% similarity to *Serratia* spp.	Accession
ChiA	Chitinase A	563	97.2% *S. ficaria*	A0A240A944
ChiB	Chitinase B	499	94.4% *S. ficaria*	A0A240C7E2
ChiA1	Chitinase A1	426	*86.2% S. nematodiphila*	A0A086GDZ3
ChiD	Chitinase D	481	89.5% *S. ficaria*	A0A240BK97
GlcNAc Binding protein A	Chitin binding		97.0% *S. ficaria*	A0A240C8H0

As previously described, lipase families from Pfam were searched on the sequences of *S. proteamaculans*. DNase plate assays on *S. proteamaculans* and *S. entomophila* have been used as species determining criteria as discussed previously. *S. entomophila* has a higher extracellular DNase response than *S. proteamaculans*, producing a larger visible halo around colonies on selective plates. The chromosome was compared for HMM's of extracellular DNases as previously undertaken with *S. entomophila* to determine any chromosomally encoded differences which would explain this

difference, or whether this is a regulatory factor. As characterised by the genus, all isolates of *Serratia* sequenced encoded extracellular *nucA* like *S. entomophila* when previously searched. When production was assessed via plate assays, a significant difference was observed, suggesting a regulatory function yet undefined in the secretion pathway. As with *nucA,* all *Serratia* isolates returned a hit for NucM- RNA endonuclease from Pfam family PF14040.

Using hmmsearch function to identify putative lipases (Table 32), 21 were found on the chromosome of AGR96X, three less than predicted in *S. entomophila* isolate 626. Searches for the same lipases families were conducted on the extended *Serratia* genus. Hmmsearch of *S. ficaria* isolate 457 revealed high degrees of amino acid similarity to both *S. entomophila* and *S. proteamaculans*. Both *S. ficaria* and *S. entomophila* have an aclacinomycin methyl esterase (RdmC), an antibiotic synthesis compound, that is not found in isolates of *S. proteamaculans*. *S. marcescens* similar to *S. proteamaculans* encodes 21 predicted lipase-associated genes, though the translated products of some of these differ. *S. marcescens* isolate 1577 encodes for two enterobactin synthase component F, in addition to phenyloxazoline synthase (*mbtB*) and a gramicidin synthase subunit, showing increased coding regions associated with antibiotic synthesis. Searches in *S. liquefaciens* chromosomes 377 and 376 again reveal similarly coded lipases. Like *S. proteamaculans*, *S. liquefaciens* again only encodes 21 predictions, with only one associated with the synthesis of an antibiotic cluster. Extracellular phospholipase A1 is uniformly predicted across all *Serratia* spp. Given their association with virulence, specific searches were undertaken for the presence of phospholipases on the chromosomes of *S. proteamaculans* isolates. Starting with AGR96X, five gene predictions were made (phospholipase, A, A1, B, and C) alongside two putative or phospholipase domain encoding genes. Comparing this to other *S. proteamaculans* isolates there were encoded differences. *S. proteamaculans* chromosomes encode phospholipase A, A1, B, and up to two copies of phospholipase C as well as a multifunctional thioesterase, but lack the putative phospholipase predicted in AGR96X.

Table 32 HMMER results for lipase family motif searches on *Serratia proteamaculans* isolate AGR96X, with percentage similarity to the most similar protein on UniProt

Protein name	Function	Size (AA)	% Similarity	Accession
Extracellular phospholipase A1	Hydrolase	320	81.3% *S. liquefaciens*	P18952
Monoacylglycerol lipase	Alpha/beta fold hydrolase	268	89.2% *S. liquefaciens*	A0A515D0X7
hypothetical protein	Lipase	615	77.8% *S. ficaria*	A0A240BYY2
Putative aminoacrylate hydrolase RutD	Alpha/beta hydrolase	335	90.1% *S. liquefaciens*	A0A515CVN6
hypothetical protein	Alpha/beta hydrolase	264	80.7% *S. liquefaciens*	A0A515CTC2
Esterase YbfF	Esterase	257	95.3% *S. liquefaciens*	A0A515CYX9
2-succinyl-6-hydroxy-2,4-cyclohexadiene-1-car	Pimeloyl-ACP methyl ester carboxylesterase	281	46.2% *Roseospirillum parvum*	A0A1G7UA85
2-succinyl-6-hydroxy-2,4-cyclohexadiene-1-car	Menaquinone biosynthesis	255	94.1% *S. proteamaculans*	A8GGZ0
2-succinyl-6-hydroxy-2,4-cyclohexadiene-1-car	Alpha/ beta hydrolase	482	96.3% *S. liquefaciens*	A0A515D232
S-formylglutathione hydrolase YeiG	Formaldeyhyde catabolic process	280	98.6% *S. liquefaciens*	A0A515CY15
Enterobactin synthase component F	Hydrolase, biosynthesis	1,233	95.1% *S. liquefaciens*	A0A515CS33
Carbonic anhydrase 1	Carbon utilisation	217	92.2% *S. liquefaciens*	A0A515CY39
hypothetical protein	Alpha/ beta hydrolase	229	93% *S. liquefaciens*	A0A515CUD9
6153-oxoadipate enol-lactonase 2	Beta-ketoadipate pathway	265	76.7% *S. ficaria*	A0A240BVC5
Carboxylesterase NlhH	Lipase 2	324	92% *S. ficaria*	A0A240C634
Putative aminoacrylate hydrolase RutD*	Nitogen utilisation and uracil catabolic process- hydrolase	267	92.8% *S. proteamaculans*	A8GCT3
hypothetical protein*	Alpha/ beta hydrolase	300	84.2% *S.liquefaciens*	A0A515CZ58
Non-heme chloroperoxidase*	Catalytic activity	273	75.7% *Sinorhizobium* sp.	A0A1B3MHF6
Acetyl esterase	Alpha/ beta hydrolase	289	90.6% *S. liquefaciens*	A0A515CVG3
N-formylmaleamate deformylase*	Alpha/ beta hydrolase	264	90.5% *S. liquefaciens*	A0A515CVL9
2-hydroxy-6-oxo-6-phenylhexa-2,4-dienoate hyd*	Alpha/ beta hydrolase	272	91.9% *S. liquefaciens*	A0A515CQ95

* Denotes predictions in hmmsearch that met the inclusion threshold for domain searches but not complete sequences.

Peptidases were again identified using the MEROP's database (full presence or absence table shown in Appendix D.6). On the assessment of the predicted peptidase output files, approximately four predicted proteins or homologs were unique to hypervirulent AGR96X. Two of these were determined to be peptidase homologues (family C44 and family M38), while a third (YhbU) was determined to be in all chromosomes, encoded on the opposing orientation. A final prediction identified an unassigned family M1 peptidase. Family M1 mainly encodes for aminopeptidases. All of these additional peptidase proteins were identifiable from MEROP's database from the *S. proteamaculans* 568 genome.

Concentrating on the chromosome of hypervirulent AGR96X, four metalloproteases were identified. Two were associated with zinc metalloprotease activity, with one extracellular metalloprotease and a putative metalloprotease. Four were also identified in other isolates of *S. proteamaculans*, as well as *S. entomophila* 626. Two extracellular serine proteases were also identified in *S. entomophila*, which was also encoded on the *S. proteamaculans* chromosome of 3041. This is less than the five predicted to be encoded in some isolates of *S. entomophila* (Chapter 5). Further analysis showed that some isolates of *S. proteamaculans* encode for more metalloprotease and extracellular serine proteases than *S. proteamaculans* AGR96X. Isolates 142 and 143 have two adjacent copies of *pmbA*, a metalloprotease found across both species, whilst MH5 has two non-co-located *ftsH* genes the products of which encode an ATP-dependent zinc metalloprotease. CfB and 20093 only encode for one extracellular serine proteases in comparison to the two found in isolates of *S. proteamaculans*, except 1048 and 1129 encode three. Overall, variability is not widely seen in the encoded metalloproteases or extracellular serine proteases of *S. proteamaculans*, with small changes across the cohort that are not associated with pathogenicity.

6.2.10 Prediction of biosynthetic clusters

AntiSMASH was again used to predict secondary metabolite clusters on the chromosomes of *S. proteamaculans* and identify any unique metabolic clusters. As with *S. entomophila*, most isolates encoded for between 7-10 secondary metabolite clusters while AGR96X itself encodes for 8 (Table 33).

Table 33 AntiSMASH results of the hypervirulent isolate AGR96X
Similarities not found in antiSMASH were reported as genomes with similar gene clusters using ClusterBlast function of antiSMASH, marked with *

Cluster	Type	Size (bp)	Similarity
1	Hserlactone	20,622	*Serratia quinivorans* NCTC13188 78%*
2	NRPS	52,595	Amonabactin P 750 *Aeromonas hydrophila* subsp. *hydrophila* 57%
3	Arylpolyene/ siderophore	40,373	83% aryl polyenes/ 77% aerobactin *Xenorhabdus*
4	NRPS	51,815	Pseudomonine *Pseudomonas fluorescens* 20%
5	Thiopeptide	26,469	O-antigen saccharide *Pseudomonas aeruginosa* 14%
6	Other	37,234	Pyrrolnitrin *Pseudomonas chlororaphis* 100%
7	NRPS-like	42,981	*S. quinivorans* NCTC13188 91%*
8	Betalactone	25,685	*Serratia* sp. BW106 95%*

Uniformity seen in *S. entomophila* for the prediction of homoserine lactone (hserlactone) and thiopeptide was not observed across the chromosomes of *S. proteamaculans* (Figure 68). These metabolites are widely described, with thiopeptides having antimicrobial properties and hserlactone as intercellular signalling molecules important in population density QS control. Four NRPS or NRPS-like metabolites were found in *S. proteamaculans* isolates like that found previously in *S. entomophila.* Using the similarity cut-offs specified in Chapter 4, most known similar clusters found in *S. proteamaculans* isolates were related to the aryl polyene siderophore cluster type, with 83% similarity to aryl polyenes from the nematode associated bacterium *Xenorhabdus doucetiae* (Grammbitter et al. 2019). This cluster similarity aligns with the results found previously referring to siderophores in *S. entomophila*, which showed 77% similarity to aerobactin siderophore from *Xenorhabdus*. Further comparisons of this region reveal that these sequences cluster in *S. proteamaculans*, with siderophore cluster and aryl polyene cluster being predicted within the same 40 Kb region (Figure 67). Predictions of a pyrrolnitrin synthetic cluster correspond to that identified in the KofamKOALA database in section 6.2.4. This region was not uniform across all *S. proteamaculans* and was not found in any *S. entomophila* isolates (Figure 68 and Table 34). All hypervirulent isolates encoded this cluster, although not unique to pathogenic isolates. The region itself showed 100% coverage and 90.75% similarity to the pyrrolnitrin cluster from *Pseudomonas chlororaphis* (MIBiG accession BGC0000924).

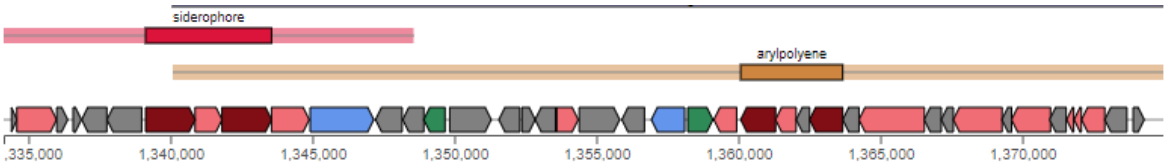


Figure 67 Schematic of predicted siderophore/aryl polyene region in *Serratia proteamaculans*
AntiSMASH prediction of siderophore aryl polyene biosynthetic cluster in most isolates of *S. proteamaculans*. When compared with *S. entomophila*, these isolates only an aerobactin siderophore show uniformity to the gene cluster in the 5' of this region. The aryl polyene is unique to *S. proteamaculans* isolates.

As shown in Table 34, no unique secondary metabolite was identified for hypervirulent AGR96X or any of the other isolates that exhibit hypervirulent pathotype (10Novel, 1129) or confer the AfpX prophage variant (20093, LC, MH5).

Table 34 AntiSMASH predicted secondary metabolite clusters in *Serratia proteamaculans*
S. liquefaciens isolates are bolded.

Cluster Type	Present in
NRPS	4 (x2), 10novel (x2), 12a (x2), 12newD (x2), 25e (x2), 142 (x2), 143 (x2), 145 (x3), 149 (x2), 299 (x2), 336 (x2), 465 (x2), 1048 (x2), 1071 (x2), 1129 (x2), 1137 (x2), 1457 (x2), 1769 (x2), 3041 (x3), 20093 (x2), AGR96X (x2), CfB (x3), **F28 (x2),** G (x2), k (x2), LC (x2), M (x2), MH5 (x2), Puna18 (x3)
Betalactone	4, 10novel, 12a, 12newD, 25e, 142, 143, 145, 149, 299, 336, 465, 1048, 1071, 1129, 1137,1457, 1769, 3041, 20093, AGR96X, CfB, **F28**, G, K, LC, M, MH5, Puna18
Thiopeptide	142, 145, 1048, 3041, AGR96X, CfB, **F28,** Puna18
Arylpolyene/ siderophore	4, 12newD, 25e, 142, 143, 145, 149, 299, 336, 465, 1048, 1071, 1129, 1137, 1457, 1769, 3041, 20093, AGR96X, CfB, **F28**, G, K, LC, M, MH5, Puna18
hserlactone	4, 12a, 12newD, 142, 145, 465, 1457, 20093, AGR96X, G, K, LC
NRPS-like	4, 10novel, 12a, 12newD, 25e, 142, 143, 145, 149, 299, 336, 465, 1048, 1071, 1129, 1137, 1457, 3041, 20093, AGR96X, **F28**, G, K, LC, M, MH5
NRPS-like thiopeptide	4, 10novel, 12a, 12newD, 25e, 143, 149, 299, 336, 465, 1071, 1129, 1137, 1457, 1769, 20093, CfB, G, K, LC, M, MH5, Puna18
Arylpolyene/ NRPS	1129, 20093, CfB
T1PKs/ NRPS	**F28**, Puna18
Pyrrolnitrin	4, 10novel, 12a, 12newD, 25e, 143, 145, 149, 299, 336, 465, 1048, 1071, 1129,1457, 1769, 20093, AGR96X, CfB, G, K, LC, M, MH5,

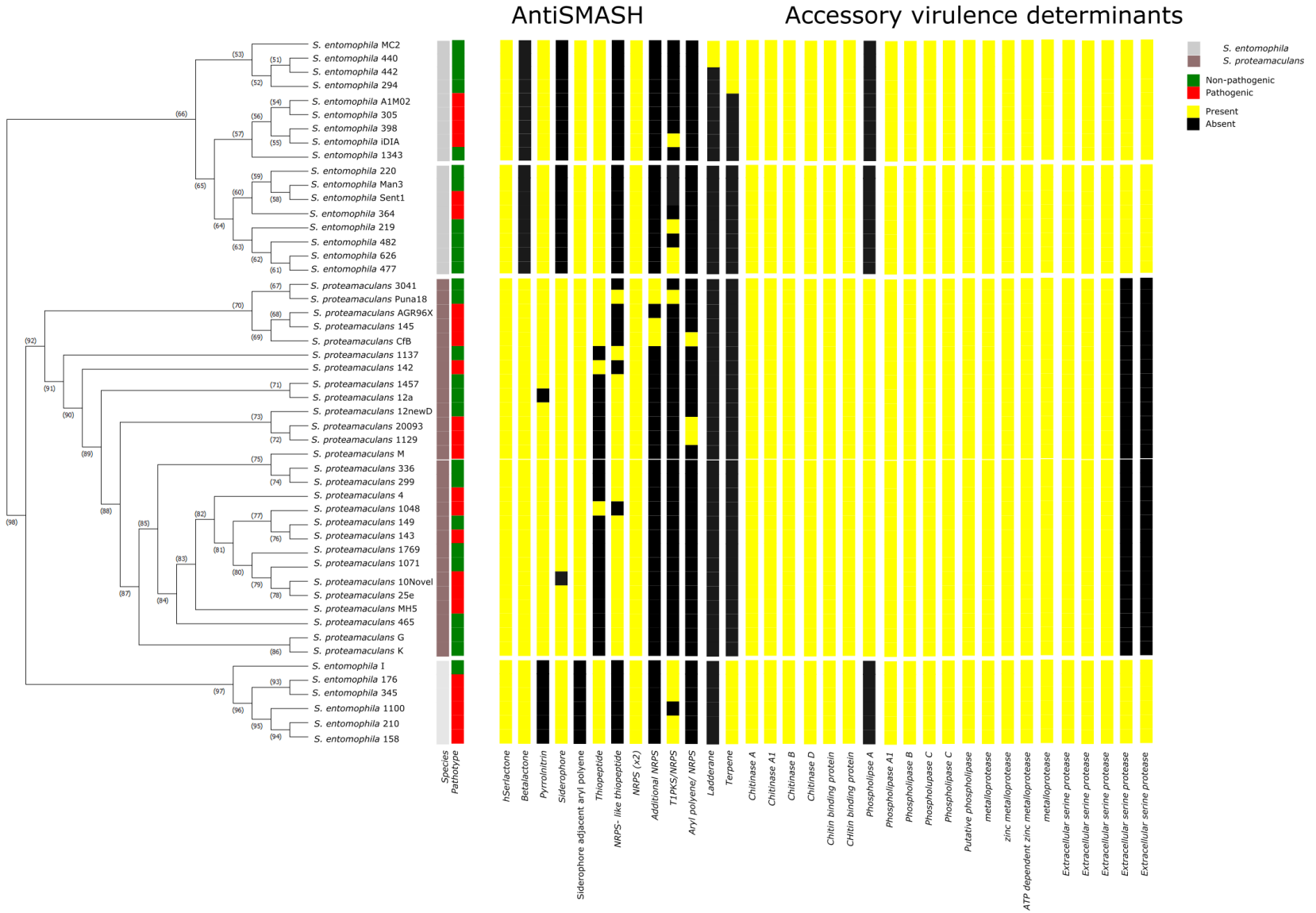


Figure 68 Presence/ absence of identifiable accessory virulence factors and biosynthetic clusters in *Serratia proteamaculans* and *Serratia entomophila*
Comparison utilises core genome phylogenies discussed in section 6.2.2 as well as incorporating a presence or absence plot based on findings in Chapter 5 section 5.2.10 and this Chapter section 6.2.9-6.2.10.

6.3 Discussion

For this discussion, the results of *S. proteamaculans* will be discussed in greater detail than *S. entomophila,* which were previously discussed in detail in Chapter 5. A full comparative analysis between *S. proteamaculans* and *S. entomophila* will be covered thoroughly in Chapter 9.

In relation to the groups defined in Chapter 3 for comparative analysis, these results show that the genomic structure of both subspecies of *S. proteamaculans* are similar in genomic makeup. Neither subspecies showed an increased level of HGT events in comparison to the other that would facilitate the inclusion of varied accessory determinants that would contribute to increased pathogenicity.

Overall, few differences are observed in encoded pathway modules for the metabolism of the *S. proteamaculans* isolates, with no indication of any difference associated with subspecies or pathotype. The starkest difference was the absence of a polyamine metabolism pathway in hypervirulent AGR96X. Polyamines have been implicated in cellular signalling regulating gene expression, survival, and stress response (Gevrekci 2017). In *Vibrio cholerae*, various polyamines tested on *V. cholerae* 01 encoded various effectors, where spermine reduced chloral toxin synthesis and various other polyamines either reduced or increased synthesis of the major pilin subunit (Goforth et al. 2013). As AGR96X has one less complete metabolic module in polyamine metabolism, it is possible that the second module present in other isolates could allow for environmental sensing to restrict the production of virulence factors much in the way demonstrated in *V. cholerae*. Analysis of whole encoded metabolic pathways revealed that *S. entomophila* in comparison to *S. proteamaculans* showed incomplete modular pathways for the synthesis of pyruvate via the De Ley Doudoroff pathway. Though all *Serratia* spp. assessed showed one core carbohydrate metabolism module for the degradation of D-galactose into pyruvate, the addition of another complete secondary pathway for synthesis in *S. proteamaculans* was interesting. Previous assessments of the itaconate degradation cluster established that the breakdown of itaconate leads to pyruvate synthesis, the same outcome is found from the De Ley Doudoroff pathway. The horizontally acquired itaconate degradation cluster (Chapter 5, Section 5.2.11) unique to *S. entomophila* isolates could indicate the adaptation of the bacterium to a more beneficial synthesis of pyruvate utilising readily available compounds synthesised from other soil-dwelling microbes.

Genomic islands (GI's) disproportionately encode for potential virulence, or survival beneficial gene clusters (Ho Sui et al. 2009). Greater diversity of GI's in various pathotypes of *S. proteamaculans* supports this theory, as this species had more predicted GI's and showed more pathotypic variation. AGR96X as the model hypervirulent isolate encodes a greater number of predicted GI's in comparison to non-pathogenic *S. proteamaculans* subsp. *quinovora* isolate 3041 and the chronic disease-causing isolate 145. In addition to GI predictions, putative prophages were identified across all *Serratia* isolates

used in the study. *S. proteamaculans* isolates have acquired a more diverse range of phage-like elements, with no predicted phage uniform across all isolates and varying degrees of predicted phage-type regions across the cohort. *S. entomophila* overall shows relatively few prophages acquired save for a few of isolates (Chatham Islands 440, 442 and 294 plus Canterbury isolate 1100) where the predicted phage account for differences that have already been noted. The similarity of phage regions and remnants in *S. entomophila* and BLASTn predictions in isolates of *S. marcescens* would imply that the prophage decay process is slow and ongoing. Defective prophage resulting from decay can result in pseudogenes (Belda et al. 2010), a possible explanation for why targeted mutagenesis of the DinI protein in Chapter 4 showed no difference in pathogenicity or MitC induction compared to the wildtype.

To relate the findings of Chapters 4 and 5 to *S. proteamaculans*, accessory virulence determinant encoding genes were identified in the hope the addition or lack thereof could attribute pathogenicity or infection to specific genes. Though identification of variation is observed, there is no trend identified to elucidate any advantageous trait to hypervirulent or chronic disease encoding isolates over non-pathogenic isolates. In fact, for some of the most pertinent accessory determinants such as chitin-binding proteins, hypervirulent isolate AGR96X did not encode the most associated genes. The majority of *S. proteamaculans* isolates assessed encoded for two chitin-binding proteins, whereby AGR96X encoded two and a small number only encoded one. In Chapter 4 it was observed that for most accessory determinants there was no statistical difference between the means of isolates of *S. proteamaculans* in relation to their pathotype or their subspecies status. These *in silico* analyses, though showing genomic level differences, support that no specific group encodes any localised beneficial trait. The inference from these results is that pathogenicity and hypervirulence are not only regulated by the presence or absence of genes but also the regulation of these genes within the cell.

In addition to the presence of an aryl polyene, *S. proteamaculans* also differ in the number and the biosynthetic clusters. *S. proteamaculans* isolates mostly encode a pyrrolnitrin synthetic cluster, which produces the diffusible antifungal antibiotic pyrrolnitrin. A significant similarity is seen between this gene cluster in *S. proteamaculans* and *Pseudomonas chlororaphis*. Functionally, mutations eliminating pyrrolnitrin production showed decreased antifungal capacity in this biocontrol agent (Selin et al. 2010). As was observed in bioassays conducted in this study, *C. giveni* larvae are susceptible to fungal pathogens of *Metarhizium* and *Beauveria,* which have another mode of action (Jackson 1990). The ability of some isolates to produce pyrrolnitrin cannot be disregarded as a potential control of fungal pathogens in the host. As a soil bacterium, it is also possible this region confers an advantage against antibiotic-producing bacteria that could pose a threat to the long-term persistence of *S. proteamaculans* in the environment.

Both *S. entomophila* and *S. proteamaculans* encode a siderophore. However, in a large number of the assessed *S. proteamaculans* isolates (all but 10novel and 12a) this region also includes the biosynthetic cluster for an aryl polyene. Both clusters have predicted similarity to those previously described in species of *Xenorhabdus*. Aryl polyenes are occasionally found in gram-negative bacteria that have host associations to any degree. Functionally, little is known about aryl polyenes and the benefits their synthesis has to organisms, though it has been proposed that they may function as protection against oxidative stress (Cimermancic et al. 2014). Previous research conducted proposed that xanthomonadins pigments in *Xanthomonas campestris* may help the bacteria evade visible light damage (Poplawsky et al. 2000). It has been suggested that aryl polyenes in gram-negative bacteria may fill this niche that is occupied by carotenoids in gram-positive bacteria which have established functionality in the reduction of oxidative stress. Clauditz et al. (2006) showed interactions between staphyloxanthin mutants in *Staphylococcus aureus* against reactive oxygen substances demonstrated the fitness benefits awarded to the wildtype strain. If this trend is continued in aryl polyenes, the ability for *S. proteamaculans* to reduce damage from oxidative stress may confer a fitness advantage in the host that S*. entomophila* does not encode for.

The encoded Aerobactin-producing siderophore is uniform across both *S. proteamaculans* and *S. entomophila*, with predictions of aerobactin consistent at approximately 77% gene-encoded similarity. Aerobactin has been identified as a critical synthetic cluster relating to host association and virulence of bacteria. In *Klebsiella pneumoniae* aerobactin has been demonstrated to increase the virulence in comparison to other siderophores produced by the bacterium (Russo et al. 2015). The interruption of aerobactin production in *K. pneumoniae* correlates with a noticeable reduction in survival of the bacterium in human sera (Russo et al. 2014). This siderophore is uniform across isolates of *S. proteamaculans* and *S. entomophila* independent of putative pathogenicity, so it is unlikely that virulence is mediated by the regulation of this siderophore. However, both pathogenic and non-pathogenic isolates are shown to survive within the host. Given the literature and uniformity of this biosynthetic cluster across all isolates it is possible that the presence of this cluster mediates survival and colonisation in the host haemolymph much like aerobactin production by the bacterium *K. pneumoniae* mediates growth in human ascite fluid.

Overall, no obvious link could be found to distinguish pathotype or subspecies apart in *S. proteamaculans*. This complements the findings in Chapter 4 where the expression of accessory virulence factors was categorised for isolates of *S. entomophila,* showing that even in *S. proteamaculans* deeper levels of disease regulation must occur to promote the pathotypic variation seen in bioassays. To further elucidate the mechanisms that underpin the hypervirulence of AGR96X,

further analysis into regulatory mechanisms of *Serratia*, and its relationship to the plasmid needs to be undertaken. The links between pathotype and isolate will be further explored in Chapter 7. Preliminary assessments of the effects of plasmids on gene regulation will then be investigated in Chapter 8.

Chapter 7

Chromosomal regulation of fitness in *Serratia*

7.1 Introduction

Regulation of virulence-associated genes has been defined by various mechanisms, including quorum sensing, global transcriptional regulators, and gene repression systems. Bacterial activation of the SOS response can also mediate the production of virulence factors (Úbeda et al. 2005). Virulence determinant of hypervirulent isolate AGR96X, the AfpX tailocin, has been reported to be inducible by initiating the SOS response with mitomycin C (MitC) (Hurst et al. 2018). This contrasts with *S. entomophila* isolate A1M02, where MitC culture supernatants showed no effect on challenged grass grub larvae, which implies that neither Afp (from chronic disease-causing pADAP plasmids), TC complex, or the *sepABC* cluster are inducible. Through expression and purification of Afp derived from an expression vector, Rybakova et al. (2013) determined that levels of over 500 particles of Afp are required to cause cessation of feeding activity. To understand the inducibility of virulence factors in other *Serratia* isolates used in this study, over three grass grub seasons MitC induction experiments were undertaken to determine any groupings or levels of inducibility to the SOS response by MitC of selected isolates. It is currently uncertain whether other amber-disease causal agents or accessory factors in other *Serratia* isolates are inducible. Further study is required to establish virulence factor induction via the SOS response in a wider range of *Serratia* isolates.

In addition to the induction of virulence factors by MitC, it was pertinent to quantify the amount of disease that could be transferred by the conjugative exchange of plasmids into differing chromosomal backgrounds. This would then resolve the hypotheses of this study that the chromosome and the plasmid are 'paired' for optimal disease expression. As outlined in Chapters 5 and 6, chromosomes of differing compositions have been identified across all isolates of *S. entomophila* and *S. proteamaculans*. To investigate the effect of chromosomal background on the fitness of the bacteria with the presence or absence of pADAP, the plasmids designated pAfpX (hypervirulent from AGR96X), p626 (chronic pADAP from isolate 626), and pADAP (chronic pADAP from A1M02, GenBank accession NC_002523.5) were conjugated into various backgrounds (Table 35). Table 35 shows the nomenclature for these transconjugants used continuously in this thesis. Nomenclature based on the addition of pADAP was chosen to reflect the origin of the plasmid e.g. pADAP-type plasmid from A1M02 was referred to as pA1M02 to reflect distinction from those from the donor chromosome of 626 and

AGR96X. The various chromosomal backgrounds were selected based on pathotype (chronic vs hypervirulent), plasmid-free non-pathogenic, and geographic isolation described in Chapter 4.

Table 35 Transconjugants used in this study

Chromosome	Wildtype Plasmid	Transconjugant	Nomenclature
iDIA	piDIA	pADAP (626)	iDIA + p626
		pAfpX	iDIA + pAfpX
477	Nil plasmid	pADAP (626)	477 + p626
		pAfpX	477 + pAfpX
		pADAP (A1M02)	iDIA + pA1M02
3041	Nil plasmid	pAfpX	3041 + pAfpX
5.6	Nil plasmid	pAfpX	5.6 + pAfpX
		pADAP (626)	5.6 + p626
		pADAP (A1M02)	477 + pA1M02

The effect of pADAP on naturally occurring non-virulent pADAP negative strains was investigated. Based on the diverse BOX PCR profiles (Chapter 4 section 4.2.3), assessment of the geographically distinct Chatham Island isolate (isolate 440) would allow for the determination of any chromosomal elements that may limit the acquisition of pADAP. The same can be predicted from West Coast isolate MC2, a plasmid-free isolate sourced from an area where grass grub does not occur, and therefore the plasmid-free state is hypothesised to be favourable. It is possible that these chromosomal backgrounds may not be suitable for harbouring a virulence plasmid and are less receptive to hosting a plasmid.

To validate that the chromosomal background affects amber disease expression, several analyses were performed enabling the potential influence of the chromosome on plasmid to be determined. These included competition assays for bacterial growth in larval isolates as well as *in vitro* growth assessments undertaken to determine whether the addition of pADAP has any effect on the fitness of the isolate. The fitness costs of standardized chromosomal background with different plasmids i.e., chronic or hypervirulent pADAP were also compared. Plasmid influence on the production of secondary virulence determinants through plate assays (outlined in section 7.2.5) was used to determine any burden on the chromosomal production of beneficial enzymes such as chitinases and proteases. Results could potentially identify whether hypervirulent plasmids are more costly on metabolism and explain why hypervirulent isolates are not readily found in field sites.

7.2 Results

7.2.1 MitC induced expression of virulence factors

MitC induced culture supernatant-derived samples were obtained following the ultracentrifuge concentration protocol utilised from Hurst et al. (2018), and described in Chapter 2, section 2.3.8. Fifty µL of each purified filtered sample was plated onto LB agar to ensure that no bacteria had made it through the purification process. Due to the temporal effects of induced virulence factors (which may decrease over 3-5 days post challenged), bioassay assessments were assessed daily over six days. This contrasts with the standard bioassay using live cells which were assessed over a 12-day duration.

Figure 69A shows that assessing live-cell bioassay data of the combined three grass grub seasons, the occurrence of disease in wildtype neat bacterial samples is consistently above 90%. These wildtype controls were used to compare the inducibility of isolates with MitC. Inducible isolates with concentrated virulence factors should show disease comparable to the wildtype isolate.

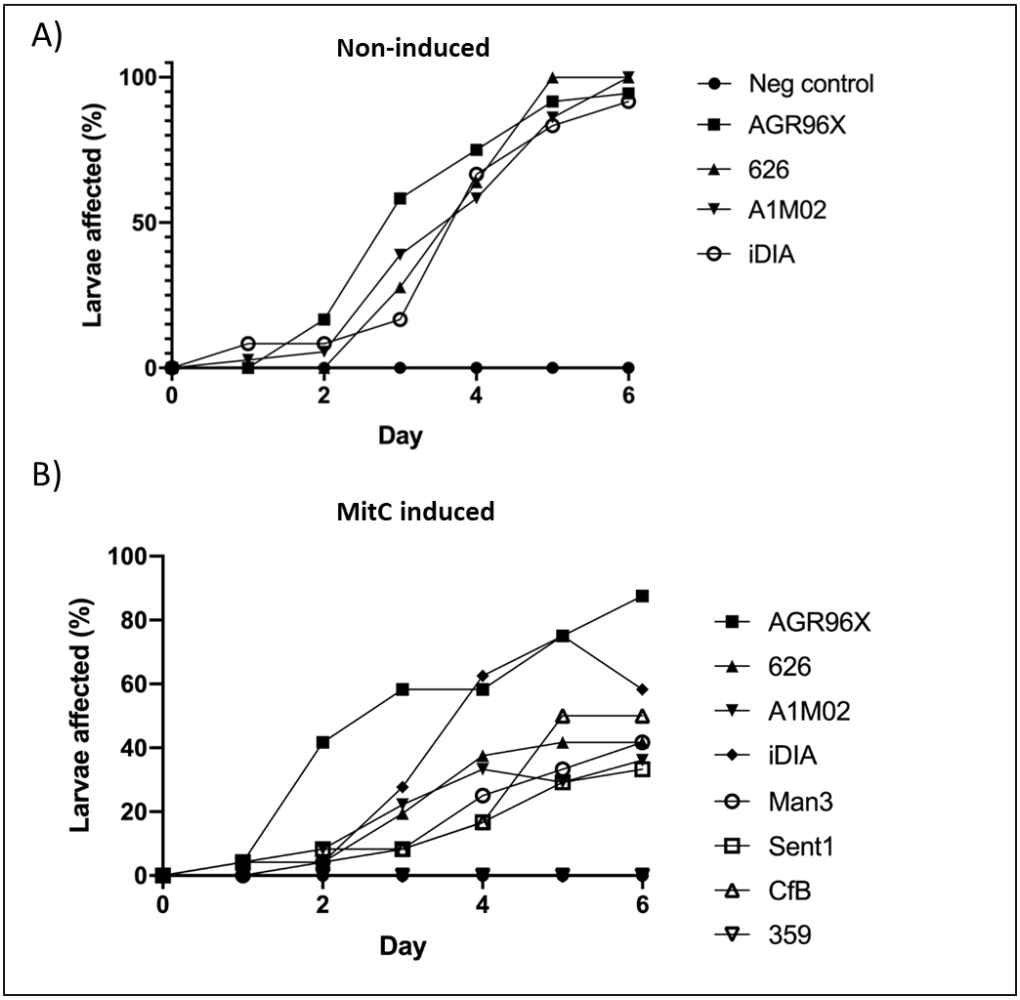


Figure 69 MitC bioassay against grass grub larvae, 2018-2020.
A) Daily progression of disease post challenge in non-induced controls (bacterial inoculation direct from plate) with the noted *S. entomophila* and *S. proteamaculans* isolates showing the percentage of observed, amber-coloured larvae six days post challenge in wildtype bacteria. B) MitC induced culture filtrates over 6 days inoculation into grass grub. Disease was based on visual changes in colour of the larvae. Non-induced control based on AGR96X.
* Indicates isolates of *Serratia proteamaculans,* other isolates included are *S. entomophila.*

Variable MitC induction effects on virulence against grass grub larvae were observed, with AGR96X consistently exhibiting higher levels of inducible response over all three seasons. Hurst et al. (2018) showed AGR96X induced AfpX resulted in a moderate degree of larval disease (64.4 ± 3.7%), which was replicated in this study (Table 36). Continuity was seen with the reduced feeding effect in AGR96X MitC induction resulting in 'nibbling' and slight maceration of the carrot cube. Both 2019 and 2020 bioassays showed the effect of MitC induced virulence factors in larvae peaked at around day four. Additionally, 2018 and 2020 assays showed that the inducibility of other isolates was relatively low, with the cessation of feeding using supernatant from isolates CfB and Man3 below AGR96X, at between 40-60% (Table 36). None of the assessed isolates exhibited a high level of disease as noted for concentrated supernatant derived from MitC induced AGR96X (Figure 69B). Of all the isolates assessed, only *S. entomophila* chronic disease-causing isolate 359 showed no inducibility (Table 36). *S. entomophila* isolate iDIA peaked in disease activity at approximately day five before disease dropping from ~80% to ~60% (Figure 69B). Stabilisation of the disease phenotype was also seen in isolates 626, A1M02, and CfB, which was not unexpected.

Table 36 Average disease and mortality rates of Mitomycin C induced isolates of *Serratia proteamaculans* and *Serratia entomophila*
Bioassays assessed on day 6, data from 2018, 2019, and 2020 were combined.
Negative control shows grass grub not dosed with any solution.
P values (Fisher's exact), significance in relation to the negative control highlighted in bold.

Inoculum source	Disease larvae (%)	± standard error (P-value)	Dead larvae (%)	± standard error (P-value)	Combined affected larvae (%)	± standard error (P-value)
Neg control	0.36	± 1.24	0	± 0	0.36	± 1.24
AGR96X	68.75	± 5.63 **(<0.001**)	14.58	± 5.14 (0.087)	83.33	± 5.43 **(<0.001**)
626*	25	± 8.83 **(0.012)**	11.11	± 5.31 (0.143)	36.11	± 8.11 **(0.001)**
A1M02*	25	± 8.83 **(0.040)**	11.11	± 5.31 (0.143)	36.11	± 8.11 **(0.001)**
iDIA*	58.33	± 6.51**(<0.001)**	2.77	± 2.77 (1.000)	61.11	± 2.77 **(<0.001)**
Man3*	20.83	± 9.11 **(0.188)**	20.83	± 8.46 **(0.050)**	41.66	± 10.27 **(0.001)**
Sent1*	29.16	± 8.56 **(0.009)**	8.33	± 5.76 (0.489)	37.49	± 10.09 **(0.004**)
CfB	50	± 7.14 **(0.003**)	0	± 0 (1.000)	50	± 15.07 **(<0.001)**
359*	0	± 0 (1.000)	0	± 0 (1.000)	0	± 0 (1.000)

* Indicates isolates of *S. entomophila*

It was noted that through the three field seasons of bioassays, induction of *S. entomophila* isolates was variable, resulting in transient effects with inducibility of *S. entomophila* isolate virulence factors (Figure 69B). This was inconsistent with previous findings of Hurst Hurst et al. (2018), where A1M02 larvae showed 3 ± 3% disease in MitC induction assays. This may reflect grass grub life stage, disease prevalence in populations, or an as yet unclassified subspecies of *C. giveni*. As shown in Figure 70, an adapted method of supernatant collection from Jackson et al. (2001) was used to assess *S. entomophila* isolate 626 and A1M02 induced samples as culture filtrates to determine if ultracentrifugation was concentrating induced virulence factors. By doing this, a small but not significant reduction in the instances of disease was observed in both isolates (Figure 70).

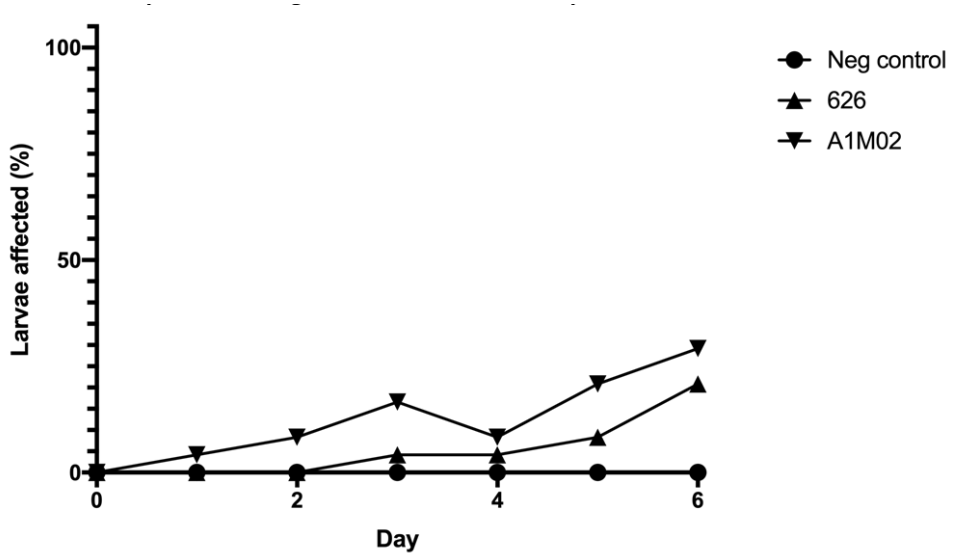


Figure 70 Percentage larvae exhibiting disease phenotype after bacterial SOS response induction with MitC using culture supernatant without ultracentrifugation over six days. Graph shows the percentage larvae exhibiting disease phenotype over six days post challenge with MitC induced supernatants. Disease was counted as visual changes in colour of the larvae. Non-induced control purified supernatant without induction of SOS response by MitC of *S. entomophila* 626.

Although virulence factors were observed to be inducible, the isolates (specifically *S. entomophila* isolates 626, A1M02, and iDIA) were inconsistent across all bioassay years. Data derived from the 2020 field season showed induction of disease at 75% of grubs assessed in AGR96X. *S. entomophila* isolates showed inducibility of disease ranging from 66% in Man3 and 626, up to 100% in A1M02 (individual year data are found in Appendix B.2). These results were not replicated over all three field seasons. In 2018, both *S. entomophila* 626 and A1M02 showed less than 20% disease in induced samples, whereas in 2019 isolate A1M02 had 0% disease by day 6 and isolate 626, 41%. Through MitC induction assays,

it can be concluded that no insect active MitC inductive factors are seen to the degree seen in the induction of AGR96X.

7.2.1.1 In silico analysis of SOS induction

To determine potential chromosomal factors relevant in the induction of AfpX during the SOS response*, in silico* searches and comparisons between isolates were undertaken. LexA-RecA mediated induction of the SOS response is controlled by the gene promotion region binding site (SOS box) that allows for the binding of LexA repressor in the absence of DNA damage. By inducing DNA damage*, recA* interrupts the binding of *lexA* to the SOS box. This leads to the upregulation of protein transcription, and therefore, as proposed with *S. proteamaculans* AGR96X, the increased production and release of AfpX (Hurst et al. 2018). To determine any influence of the SOS box on altered gene expression, the genome sequences of each of the seven isolates used in MitC assays were assessed for the presence of putative SOS box binding sites. The SOS binding sites sequence motif for gammaproteobacteria, CTGT(N_8)ACAG, described by Erill et al. (2003), was assessed using the prosite motif search function in Geneious. Known *lexA* regulated genes in *E. coli* were then used to strengthen the *in silico* results (Maslowska et al. 2019). A full list of predicted sites and the associated downstream gene can be found in Table 38 for AGR96X. All SOS binding sites of the selected isolates are listed in Appendix E.1.

Table 37 Motif search results of CTGT(N_8)ACAG on various genomes of *Serratia* isolates used in MitC induction.

Motif searches allowed for no mismatches on the eight defined nucleotides.
* indicates isolates of *S. entomophila*
[1] indicates consistently inducible virulence factors with MitC

Isolate	SOS BOX motif hits
AGR96X[1]	61
626*	48
Man3*	44
A1M02*	43
iDIA*	43
CfB	61
Sent1*	42

Chromosomal searches revealed AGR96X and CfB has a significantly higher number of hits compared to isolates of *S. entomophila* (Table 37). Comparatively, the SOS box hits in the genomes of other *S.*

proteamaculans were lower. The reduction of response between AGR96X and CfB from 2020 is not enough to rule out the potential that CfB, to an extent, is also inducible. Searches between predicted motif sites were focused on differences between inducible AGR96X and non-inducible *S. proteamaculans* and *S. entomophila* isolates (Table 37). Full annotations of predicted motif sites can be found in Appendix E.1

S. proteamaculans isolates assessed predicted an SOS binding site on a *fusAB*-like tetrapartite efflux system operon with 94.8% similarity to FUSC family protein in *S. liquefaciens*. This region functions as a substrate-specific transmembrane transporter. This operon is present in duplicate copies in *S. proteamaculans*, with one operon having a predicted SOS binding site within the *fusC* gene, where this operon is not found in *S. entomophila.* SOS box predictions of AGR96X revealed five transcriptional regulators encoding a down-stream putative SOS binding site. Of the five transcriptional regulators, three LysR family regulators were identified. Adjacent potential regulatory genes were established as a transport protein and a putative pyrimidine utilisation gene. The third was predicted as hypothetical. BLASTp displayed this protein as 'uncharacterised' with 53.3% AA similarity to an uncharacterised protein of *S. ficaria* (Table 38).

Table 38 Predicted genes associated with putative SOS binding sites in *Serratia proteamaculans* AGR96X

SOS binding sites were identified using the gammaproteobacterial motif CTGT(N_8)ACAG and the Geneious motif search function. Sequence position refers to the position of the putative SOS box to the closest downstream CDS. Predicted hypotheticals are defined with BLASTp and the closest similarity is shown.

Gene	Function	Sequence position
RecN*	DNA repair	26 bp 5'
LexA*	SOS repressor	5 bp 5'
LexA*	SOS repressor	28 bp 5'
DsrB*	Putative	266 bp 5'
Multidrug resistance protein	Transporter	5 bp 5'
hypothetical protein	Uncharacterised protein *Serratia ficaria* 82.2%	271 bp 5'
hypothetical protein	Uncharacterised protein 43.6% *Klebsiella pneumoniae* IS39	129 bp 5'
Phage tail length tape-measure protein 1		Internal
Formiminoglutamic iminohydrolase	hydrolase, carbon-nitrogen	Internal
RecA *		77 bp 5'
QnrB	Pentapeptide repeat protein	10 bp 5'
2-C-methyl-D-erythritol 2,4-cyclodiphosphate synthase		Internal
Nitrate ABC transporter, ATP-binding protein	permease	Internal
hypothetical protein	hypothetical 2	Internal

YphG	Uncharacterised protein	Internal
Dihydroxy-acid dehydratase		Internal
YebG*	DNA damage-inducible gene in SOS regulon, expressed in stationary phase	23 bp 5'
Allophanate hydrolase 2 subunit 1	Hydrolase	87 bp 5'
2,5-dioxovalerate dehydrogenase (EC 1.2.1.26)		Internal
hypothetical protein	35.6%Unchracterised protein *Serratia odorifera*	Internal
Nuclease precursor (EC 3.1.30.2)		30 bp 5'
Cell division inhibitor		21 bp 5'
3-oxoacyl-[acyl-carrier-protein] synthase II		Internal
23S rRNA	(guanine(2445)-N(2))-methyltransferase (EC 2.1.1.173)	Internal
Hypothetical	Outer membrane autotransporter metallopeptidase activity 46.6% *Serratia fonticola*	Internal
Hypothetical	Uncharacterised protein 43.3% *Marinomonas pollencensis*	75 bp 5'
Putative tail fibre protein^{o}		132 bp 5'
Intracellular protease	Protease	Internal
hypothetical protein^{o}	Putative tail spike protein 42.2% *Serratia* phage eta	39 bp 5’
Putative kinase		Internal
Hypothetical	Phage tail-like protein 33.6% Enterobacter sp.	Internal
Hypothetical	Uncharacterised protein 84.4% *Serratia odorifera*	24 bp 5’
YcbJ	Uncharacterised protein	Internal
Possible abortive infection phage resistance protein		Internal
AlkB	Alpha-ketoglutarate-dependent dioxygenase	10 bp 5'
Lactoylglutathione lyase and related lyases		17 bp 5'
UmuD*	Error-prone repair protein	93 bp 5'
FtsK*	DNA translocase	65 bp 5'
Tetrapartite efflux system		Internal
LysR[1]	Transcriptional regulator	102 bp 5'
AcrR family	Transcriptional regulator	Internal
LivK	High-affinity leucine-specific transport system	Internal
DNA-3-methyladenine glycosylase	glycosylase	14 bp 5'
YjcS	Putative alkyl/aryl-sulfatase	Internal
Hypothetical	Fimbrial anchor 53.7% *Erwinia tasmaniensis*	Internal
Uncharacterized protein YtfB	Uncharacterised protein	Internal
DNA polymerase IV		16 bp 5'
Glutamine amidotransferase, class I		Internal
Permease, YjgP/YjgQ family	Permease	Internal
PpiA	Peptidyl-prolyl cis-trans isomerase precursor	Internal
Transposase		Internal
ATP synthase protein I	ATP synthase	Internal

Transcriptional regulator	GntR family domain / Aspartate aminotransferase	Internal
Transcriptional regulator	GntR family domain / Aspartate aminotransferase	Internal
Ribulosamine/erythrulosamine 3-kinase	potentially involved in protein deglycation	Internal
DNA-directed RNA polymerase beta subunit		295 bp 5'
HypF	hydrogenase metallocenter assembly protein	Internal
tRNA-Val-GAC		13 bp 5'
fusBC-like	Tetrapartite efflux system, inner membrane component	Internal
LysR[1]	Transcriptional regulator	76 bp 5’
LysR[1]	Transcriptional regulator	9 bp 5’

* Denotes SOS binding sites previously identified in *Escherichia coli*

[1] Indicates SOS binding sites unique to AGR96X when compared to other assessed isolates from Table 37

[o] Indicates genes associated with a predicted phage region

The presence of predicted SOS binding sites in AGR96X was also compared to the presence of phages predicted in Chapter 6. It is conceivable that the presence of any horizontally acquired DNA under regulation by the SOS response conferred some benefit to the overproduction of AfpX observed in induced AGR96X. Of the four phages predicted in AGR96X, only two genes associated (located on contig one) have predicted SOS binding sites (Table 38). In addition to phage-encoded proteins, a putative AGR96X phage-associated gene was identified to contain a predicted SOS binding site. A predicted abortive infection phage resistance protein (AIPR) on the chromosome of AGR96X was identified. BLASTp and Uniprot blast function identified this protein as having 76.6% amino acid identity to an AIPR family protein from *Pectobacterium odoriferum.*

In addition to LexA-RecA dependent SOS response induction, it is also possible that a LexA-independent SOS response as described in organisms such as mycobacteria (Müller et al. 2018) mediates the over-production of virulence factors in AGR96X. Further analysis needs to be undertaken to fully understand the mechanism of induction in AGR96X. Work by Hurst (unpublished) seeks to unravel the mechanism of this induction in AGR96X compared to isolates of *S. entomophila* utilising transcriptome data of induced samples. Identified regions could allude to target areas for mutagenesis to identify the potential mechanism on the chromosome that regulates the induction of potential virulence factors under stress conditions. Further conclusions could potentially be drawn from additional MitC inductions of other AfpX encoding *S. proteamaculans* isolates such as MH5, which exhibits variable pathogenicity.

7.2.2 Conjugation of plasmid types into various chromosomal background

Building on a previous study by Sitter (2020), together with observations made in this thesis (Chapters 4 and 6), reasons why the hypervirulent pathotype is only found in *S. proteamaculans* isolates were explored. While previous research by Grkovic et al. (1995) deduced that the ability to conjugate megaplasmids from *S. entomophila* into other species of *Serratia* was possible, the degree of pathogenicity conferred and how comparable to the wildtype disease has yet to be defined. By defining the transfer of pathogenicity, the impact of the acquired plasmid on growth, infectivity, and expression of secondary accessory factors could then be described (described in Chapter 4, section 4.2.5). In this section, isolates of both *S. entomophila* and *S. proteamaculans,* selected based on pathotype and geographic isolation (Table 39), were tagged allowing selection as a recipient of donor plasmids (Table 40). Primarily, various *S. entomophila* isolates were conjugated with the hypervirulent pAfpX to determine whether the full breadth of the pathotype was acquired and whether any negative effects were observed on host cell metabolism.

Table 39 Isolates with varying pathotypes and plasmids encoded selected for tarns conjugant experiment

Species		Isolate	Resident Plasmid	Putative pathotype	Location (NZ)
Serratia	*entomophila*	MC2*	None	Non-path	West Coast
Serratia	*entomophila*	440*	120 Kb	Non-path	Chatham Islands
Serratia	*entomophila*	iDIA	pADAP	Chronic	Canterbury
Serratia	*entomophila*	477	None	Non-path	Canterbury
Serratia	*entomophila*	5.6	None	Non-path	Canterbury
Serratia	*proteamaculans*	3041	None	Non-path	Manawatū
Serratia	*proteamaculans*	AGR96X*	pAfpX	Hypervirulent	Canterbury

* Denotes isolates where transconjugants were not successfully produced.

To test this, backbone tagged plasmids from *S. entomophila* isolates A1M02, 626, and *S. proteamaculans* AGR96X were used in conjunction with chromosomally tagged 5.6, 477, 3041, and plasmid cured iDIA. The full list of conjugates used in this study are listed in Table 40.

Table 40 Successfully transconjugated bacterial chromosomes with the inserted plasmid and the corresponding chromosomal and plasmid markers

Species	Isolate	Chromosomal marker	Plasmid	Plasmid backbone marker
Serratia proteamaculans	3041	Spec	pAfpX	Cm90
Serratia entomophila	477	Tet	pAfpX	Cm90
			pADAP	Kan100
			p626	Cm90
Serratia entomophila	iDIA	Tet	pAfpX	Cm90
			p626	Cm90
Serratia entomophila	5.6	Spec	pAfpX	Cm90
			pADAP	Kan100
			p626	Cm90

Transconjugants were validated using BOX-PCR, Kado and Liu (1981) megaplasmid visualisation, *repA* targeted PCR and AfpX18 targeted PCR (Chapter 2, section 34). Examples of megaplasmid visualisation gel and BOX-PCR gel are shown in Figure 71. Appropriate banding was found to confirm conjugation of plasmid types into chromosomal backgrounds in panel A. Panel B shows BOX-PCR bacterial fingerprinting also show similar profiles between the transconjugants and their wildtype. Although there is differential banding, based on linking these results with Chapter 4 (Figure 13), there doesn't appear to be a relationship between binding and the presence/absence of pADAP or its variants. As explained by Bilung et al. (2018), the effectiveness of BOXA1R bacterial fingerprinting relies on it being used as a complementary typing tool.

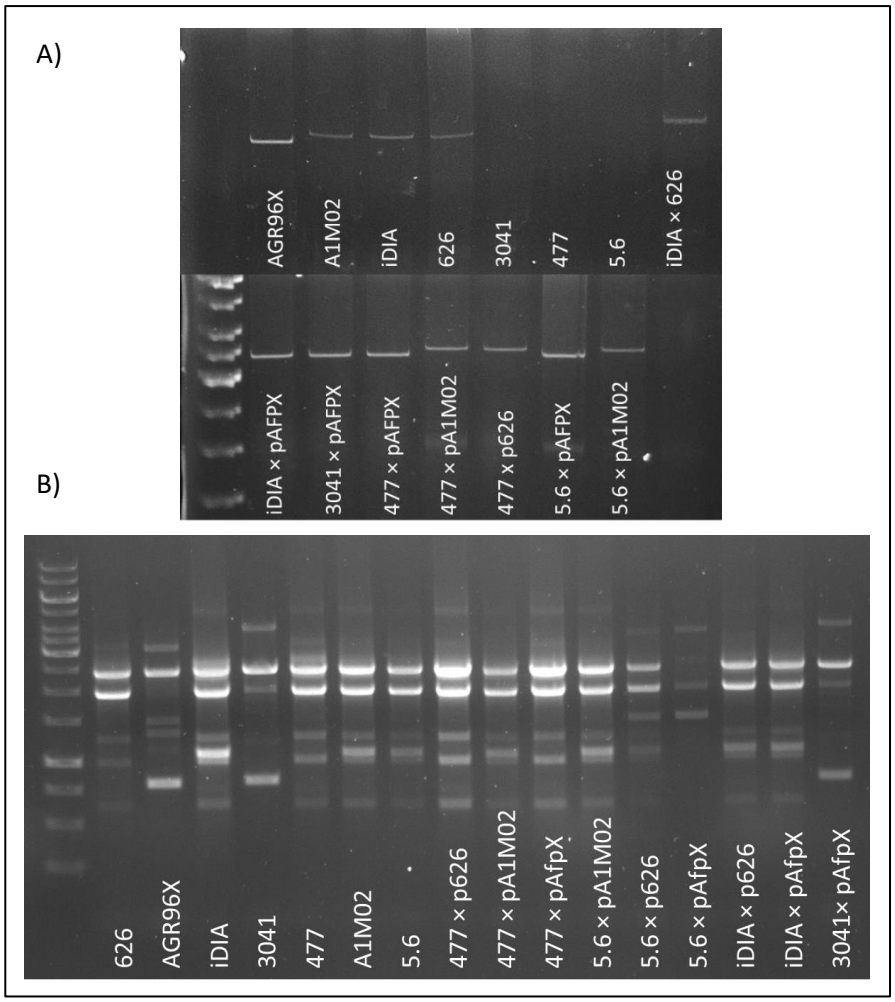


Figure 71 Kado and Liu megaplasmid visualisation and BOX PCR validation of transconjugants
Isolates shown alongside their wildtype controls.
A) shows megaplasmid visualisation of transconjugants and wildtypes B) shows BOX bacterial fingerprinting profile.

Attempts to conjugate the plasmid variants (pADAP from 626 or A1M02 and pAfpX) to either the naturally pADAP-free 440 (but has a 120 Kb plasmid) and plasmid-free isolates MC2, 3041, and the AfpX bearing AGR96X were unsuccessful. Either no, or only a subset of plasmids in these instances were able to be conjugated (Table 40). In the attempt to construct variable transconjugants of different backgrounds, heat curing of 440 (Chatham Islands) and AGR96X and its derivative XΔNOVA were unsuccessful. XΔNOVA and 440 were subjected to several further methods to cure plasmids i.e., heat curing or use of stress-induced by detergents as described in Chapter 2 section 2.3.5. These strategies were also unsuccessful. In addition to these methods forced conjugations were also attempted between XΔNOVA and tagged A1M02, where no plasmid transfer was noted. It is postulated that the inability to plasmid-cure these isolates could be due to tight regulation between

the plasmid and chromosome where the addition of pADAP variant to AGR96X/ XΔNOVA would serve as less beneficial to the cell than retaining the host plasmid, or via a plasmid handcuffing mechanism.

Based on this only a select set of chromosome backgrounds (listed in Table 40) were assessed in this study. Isolates selected were from varying backgrounds and pathotype. One of the main interests was to look at the effect that the acquisition of plasmids has on pathogenicity, by natural non-plasmid bearing isolates *S. proteamaculans* 3041, and *S. entomophila* 477. Chromosomes of *S. entomophila* MC2 and 440 were successfully tagged. However, no successful transconjugants or these isolates were acquired, even when varying conjugation conditions and concentration ratios of the donor to recipient. The inability to conjugate plasmids into MC2 could support the hypothesis that at least some naturally plasmid-free isolates may not be able to retain a plasmid in the environment.

7.2.3 LC_{50} and LT_{50} against grass grub larvae

Bioassays to calculate LC_{50} and LT_{50} were performed to allow a more accurate reflection of plasmid acquisition on pathogenicity. Due to high background levels of disease in larvae collected at AgResearch, Lincoln (30-50% on pre-feeding) in 2020, and the delayed season start due to unforeseen international SARS-CoV-2 lockdowns, a margin of error was allowed for blank thresholds accepted. disease cut-off threshold for background disease was 25% in the negative control. Additionally, where mortality in AGR96X wasn't as high in comparison to previous seasons (though infectivity remained consistently high) there was the potential for this to be attributed to localised immunities (genetic diversity) developed in collection sites. In general, larvae in the 2020 season were delayed in growth, likely due to a prolonged dry summer, in turn affecting metabolism. Disease phenotype classifications were assigned based on the relative disease and percentage dead inflicted on larvae inoculated with transconjugant isolates compared to the plasmid-free controls and wildtype original plasmid donor isolate (Table 41). While AGR96X mortality was below 75% (hypervirulent is defined as mortality above 75%) the mortality rate remained the highest, therefore was used within the previous context definition as hypervirulent. The LC_{50} bioassays were conducted in triplicate to further solidify the findings in initial bioassays, where slightly elevated levels of background disease were reported. Where the disease was observed to be acquired, but in separate bioassays was inconsistent in its effect, the result was classified as "variable pathotype".

Wildtype LC_{50} bioassays conducted on pathogenic *S. entomophila* isolates 626, iDIA and A1M02 showed consistency in the disease onset at approximately 3-4 days (Figure 72). The diseased larvae were amber in colouration and paralysis of the mandibles resulted in a complete cessation of feeding activity. At high concentrations, this was uniform across all individuals. iDIA had a higher rate of

mortality compared to other *S. entomophila* isolates (8.1×10^7 CFU inoculant). For AfpX encoding AGR96X (4.9×10^7 CFU inoculant), the disease phenotype displayed a broader scope of disease traits. As described previously by Hurst et al. (2018), cessation of feeding was not observed with hypervirulent AfpX and, instead, a slight maceration effect on the edges of the carrot was found. Full cessation of feeding is, however, seen in *S. entomophila* Afp encoding isolates. Amber discolouration was less uniformly seen in *S. proteamaculans* AGR96X, although the cessation of feeding was noticeable in 1-2 days post-inoculation. A high level of larval mortality was also noted with inoculation of AGR96X. Non-plasmid bearing isolates *S. entomophila* 477 and 5.6, alongside *S. proteamaculans* 3041 showed no difference to the untreated blank control, confirming for non-pathogenic isolates (Table 41).

For transconjugants of either pADAP from A1M02 (2.3×10^8 cells average inoculant) or 626 (5.1×10^8 cells average inoculant), the recipient isolate caused more uniform disease. With the transconjugant 5.6 + pA1M02 challenged larvae exhibited a disease pathotype after approximately four days, whilst the addition of pADAP from 626 (5.6 + p626) caused amber discolouration but a non-uniform cessation of feeding (Table 41) (Figure 72 and Figure 73). Restoring pADAP (from A1M02) to heat-cured isolate 5.6, showed disease and larval death do occur but this was not consistent, and amber discolouration was less pronounced than with the addition of pADAP from 626. Conjugation of the hypervirulent plasmid from AGR96X (5.6 + pAfpX) showed the same phenotypic symptoms as associated with AGR96X, i.e., 'nibbling 'of the carrot, but did not cause observable disease until day three. From day four 5.6 + pAfpX showed more uniform amber colouration than infection by the wildtype AGR96X. Overall, the mortality rate due to *S. entomophila* 5.6 + pAfpX was greatly reduced, resulting in ~30% less mortality than wildtype AGR96X (52.38 ± 11.16%). There was no statistical difference (P=0.224) between the transconjugant 5.6 + pAfpX and the blank control.

Wildtype 477, as shown in Table 41 and previous bioassays in Chapter 4 section 4.2.2, is a non-pathogenic, pADAP-free isolate. By 3 days post-challenge of *C. giveni* larvae the *S. entomophila* 477 + p626 exhibited disease onset with consistent discolouration and cessation of feeding occurring across all three bioassays similar to that observed for WT *S. entomophila* 626 (Figure 72 and Figure 73). In one bioassay all larvae remained healthy until day 4, again not exhibiting the hyper-disease observed for AGR96X. Using 477 + pAfpX, the LC_{50} was $6.8 \times 10^{-6} \pm 6.8 \times 10^{-6}$ cells which were on average 100 fold higher than that of wildtype AGR96X ($4.9 \times 10^{-4} \pm 3.6 \times 10^{-4}$ cells). Using 5.6 + pAfpX the LC_{50} was similar to 477 + pAfpX ($1.08 \times 10^{-7} \pm 1.01 \times 10^{-7}$ cells) (Table 41). Typically, with AGR96X, mortality is in the range of 75% within 5-12 days (as defined for hypervirulence) whereas with these pAfpX

transconjugants mortality was much reduced (Table 41). *S. entomophila* isolate 477 with conjugant plasmids from either 626 or A1M02 exhibited similar disease levels to the wildtype donor.

Wildtype iDIA causes a more rapid disease progression than 626 or A1M02, resulting in higher levels of mortality. The addition of p626 into the plasmid-cured iDIA chromosome restored the pathotype by day 4 (Figure 72 and Figure 73). Consistently, the transconjugant iDIA + p626 exhibited a similar virulence profile to wildtype IDIA, showing the addition of the 626 derived pADAP does not affect pathotype. Interestingly, incorporating pAfpX into iDIA was the most responsive of all the *S. entomophila* transconjugants, showing comparable LC_{50} to that of *S. proteamaculans* 3041 + pAfpX (Table 41).

The addition of pAfpX into non-pAfpX bearing chromosomes i.e., 5.6, 3041, and 477, resulted in some inconsistencies. A similar delayed disease progression was observed in 3041 + pAfpX, with an LT_{50} of seven days. When comparing the wildtype plasmid-free isolates of *S. entomophila* 477 and *S. proteamaculans* 3041 pathogenic ability was acquired. The level of virulence however was not at the same level as the wildtype plasmid donor, especially notable for those conjugates bearing the pAfpX plasmid. The LC_{50} of AGR96X suggests that the iDIA and the *S. proteamaculans* 3041 chromosomes, unlike *S. entomophila* isolates 5.6 and 477, encodes attributes that can complement pAfpX fully (to be assessed further in section 7.2.7).

Table 41 Day 12 percentage disease based bioassay data for all transconjugants and the wildtype controls, including LC_{50} and LT_{50} with standard error calculation
P values (Fisher's exact), with significance in relation to the negative control, highlighted in bold.

Isolate	LC_{50}[1] CFU/ml ± standard error			LT_{50}[2] (day)	Diseased (%)	± standard error (P-value)	Dead (%)	± standard error (P-value)	Combined (%)	± standard error (P-value)	Transconjugant pathotype
Blank		-		-	0.00	±0	30.56	±7.78	30.56	±7.78	
5.6		-		-	28.57	±10.10 **(0.001)**	9.52	±6.56 (0.103)	38.10	±10.85 (0.575)	Non-path
477		-		-	5.00	±5 (0.357)	15.00	±8.19 (0.334)	20.00	±9.17 (1.000)	Non-path
3041		-		-	0.00	±0 (1.000)	33.33	±8.33 (1.000)	30.30	±8.12 (1.000)	Non-path
626	2.6×10^{-5}	±	1.9×10^{-5}	4	71.43	±10.10 **(<0.001)**	28.57	±10.10 (1.000)	100.00	±0 **(<0.001)**	Pathogenic
A1M02	2.9×10^{-5}	±	2.7×10^{-5}	3	71.43	±10.10 **(<0.001)**	28.57	±10.10 (1.000)	100.00	±0 **(<0.001)**	Pathogenic
AGR96X	4.9×10^{-4}	±	3.6×10^{-4}	3	42.86	±11.06 **(<0.001)**	52.38	±11.16 (0.158)	95.24	±4.76 **(<0.001)**	Pathogenic
iDIA	2.0×10^{-4}	±	1.5×10^{-2}	4	62.50	±8.69 **(<0.001)**	28.13	±8.07 (1.000)	90.63	±5.23 **(<0.001)**	Pathogenic
3041 + pAfpX	1.9×10^{-6}	±	1.5×10^{-6}	7	36.36	±8.50 **(<0.001)**	33.33	±8.33 (1.000)	69.70	±8.12 **(0.002)**	Pathogenic
477 + p626*	8.6×10^{-5}	±	8.5×10^{-5}	4	72.73	±7.87 **(<0.001)**	27.27	±7.87 (0.797)	100.00	±0 **(<0.001)**	Pathogenic
477 + pA1M02*	4.9×10^{-5}	±	2.4×10^{-5}	4	54.55	±8.80 **(<0.001)**	36.36	±8.50 (0.621)	90.91	±5.08 **(<0.001)**	Pathogenic
477 + pAfpX	6.8×10^{-6}	±	6.8×10^{-6}	4	60.61	±8.63 **(<0.001)**	21.21	±7.22 (0.422)	81.82	±6.81 **(<0.001)**	Variable
5.6 + p626	5.3×10^{-7}	±	5.3×10^{-7}	4	28.57	±10.10 **(0.001)**	33.33	±10.54 (1.000)	61.90	±10.85 **(0.028)**	Variable
5.6 + pA1M02	6.4×10^{-6}	±	6.4×10^{-6}	4	72.73	±7.87 **(<0.001)**	27.27	±7.87 (0.797)	100.00	±0 **(<0.001)**	Pathogenic
5.6 + pAfpX	1.1×10^{-7}	±	1.01×10^{-7}	3	27.27	±7.87 **(0.001)**	18.18	±6.81 (0.274)	45.45	±8.80 (0.224)	Variable
iDIA + p626*	2.2×10^{-4}	±	1.1×10^{-4}	4	78.79	±7.22 **(<0.001)**	21.21	±7.22 (0.422)	100.00	±0 **(<0.001)**	Pathogenic
iDIA + pAfpX	7.7×10^{-5}	±	7.1×10^{-5}	7	36.36	±8.50 **(<0.001)**	12.12	±5.76 (0.083)	48.48	±8.83 (0.146)	Variable

* Denotes where a transconjugant shows proximity to the LC_{50} of the wildtype plasmid equivalent
[1] Undertaken via Probit analysis
[2] Statistical Survival analysis

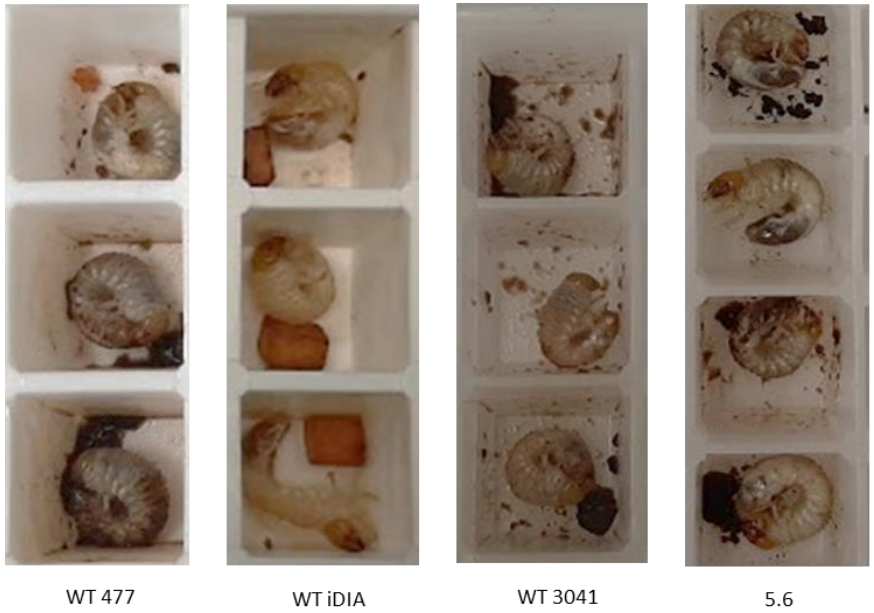


Figure 72 Sample bioassay disease instances of wildtype isolates at day six for wildtype controls
Photographs of representative phenotypes of disease exhibited by using each wildtype isolate. Cessation of feeding is noted where untouched carrots were observed. Amber discoloration is indicative of amber disease phenotype (seen strongly in WT iDIA). Isolates 477, 3041 and 5.6 are naturally plasmid free non-pathogenic. iDIA is a chronic pathogen.

Figure 73 Sample bioassay disease instances of transconjugants at day seven.
Photographs of representative phenotype of disease exhibited after treatment with plasmid transconjugant isolates. Cessation of feeding noted where untouched carrots were observed. Amber discoloration indicative of amber disease phenotype.

7.2.4 Bacterial infectivity

To further complement the data acquired during the LC_{50} and LT_{50} bioassays and try and determine the mechanism behind the observed aberrant phenotypes of the various transconjugants, an infectivity assay coupled with bacterial re-isolation was conducted. Each assessment was undertaken in triplicate to determine an average CFU for each isolate from the host macerate sample every three days. The transconjugant was then selected on LB agar with selective antibiotics for the chromosome and plasmid tags, against control for the wildtype variant. Day 12 samples were patched back onto plates with antibiotics selective for the plasmid tag, to determine whether there was any plasmid loss (as described in Chapter 2, section 2.6.4).

By day 12, similar CFU's were isolated from larval macerates in 3041 + pAfpX compared to the wildtype non-pathogenic control, 3041. When comparing across the four time points, the 3041 transconjugant did not deviate significantly from the wildtype control (P=0.810). For 477 and its AfpX conjugant, differences in cell numbers were observed. A minimum log-fold decrease in the average re-isolation amount of 477 + pAfpX was observed over the 12. Relative to the wildtype 477 control (non-pathogenic) low cell numbers observed on all-time points, implying a reduced infectivity rate when a plasmid is introduced. Statistical analysis of the data transformed to normalised starting CFU (Figure 74) showed that the difference of re-isolation CFU between wildtype 477 and its plasmid transconjugants was not significant (all three P>0.05). While the wildtype doesn't have a significant advantage at host colonisation (expected as a non-pathogenic isolate), 477 + p626 trends lower than the other three, which cluster quite close together at day 12 (Figure 74).

Observing the bioassay data for iDIA the trend is that the WT iDIA has higher concentrations at day 12 than either of the plasmid transconjugants, with a one log fold difference for iDIA + p626. Statistical analysis of the corrected data showed that this was not significant (p626 p=0.081 and pAfpX p=0.465). Between A1M02 as a WT control and the transconjugants of isolate 5.6, more consistency in the assay was observed, with day 15 re-isolation CFU's observed within a log fold difference. This was expected due to the native plasmid (pADAP) of isolate 5.6 being restored. Again, no statistical difference was observed.

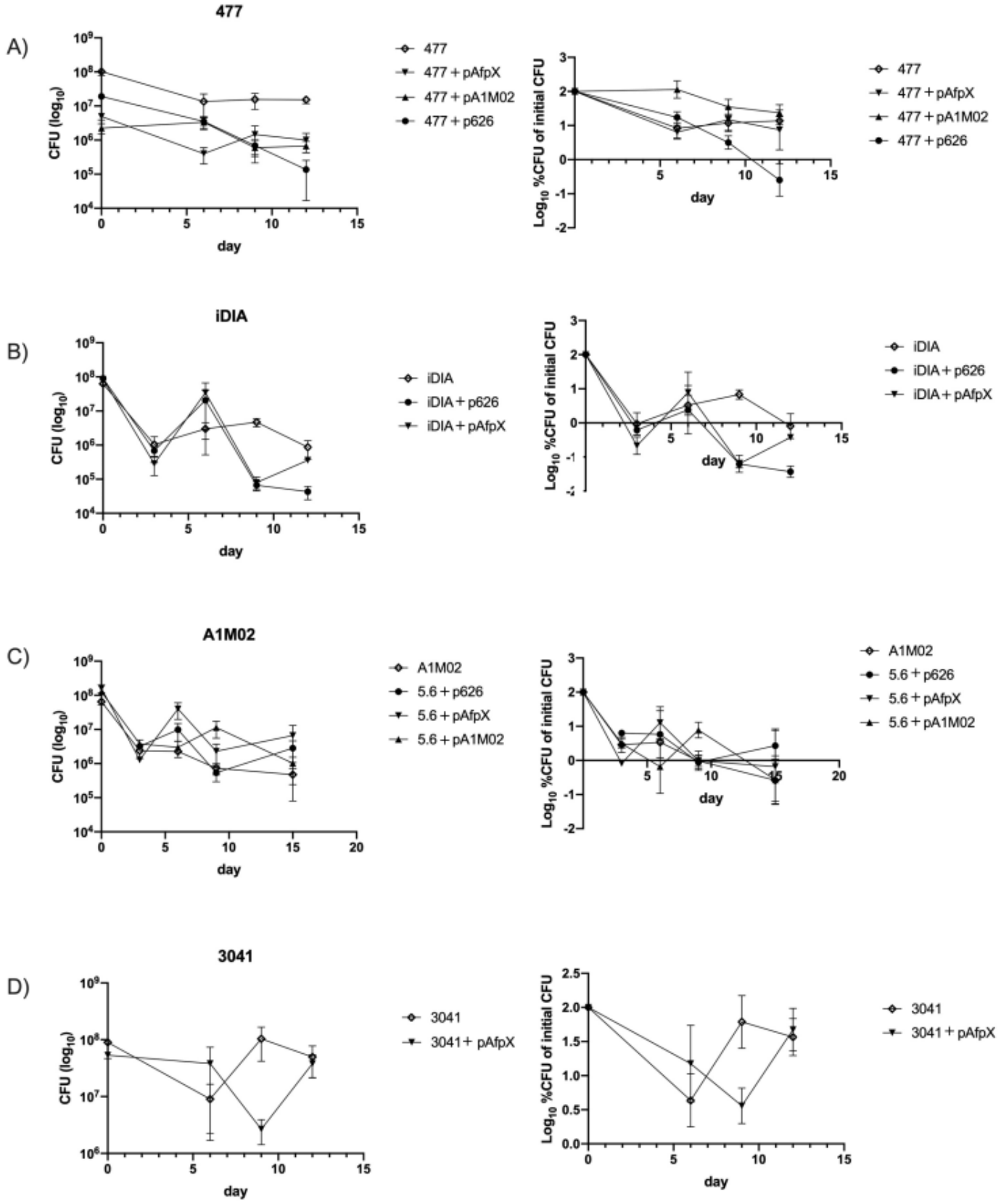


Figure 74 Number of CFUs recovered from host macerates of larvae inoculated with transconjugants in relation to their relative wildtype strains over 15 days and normalised data represented as log_{10} transformed percentage of initial CFU.

A) plasmid transconjugants into the WT *Serratia entomophila* 477 chromosome B) plasmid transconjugants and WT *S. entomophila* iDIA, C) *S. proteamaculans* WT 3041 and its pAfpX transconjugant and D) *Serratia entomophila* WT A1M02 and its three plasmid transconjugants. All wildtype strains are shown as white placements. Standard error bars are shown. All normalised data graphs shown to the right of the raw bioassay data.

In assessments of the isolates 477 + pAfpX (Tet, Cm resistant) and 3041 + pAfpX (Spec, Cm), almost no colonies could be isolated from the initial day 12 dilutions onto plasmid selective Cm medium. To investigate whether plasmid loss had occurred colonies from the chromosomal tag-specific antibiotic plate (Tet or Spec) were patched onto plasmid selective Cm plates where 100% growth was observed. Relating this to previously discussed results, there is the potential that the stress of outcompeting other bacteria and direct patching may have stunted growth in the presence of the antibiotic, resulting in a low colony yield. This would be supported by the smaller colonies observed on agar plates (without antibiotics) for both of these transconjugants (Figure 75 and Figure 76).

To conclude, plasmid loss was not observed over a twelve-day assessment period, though overall re-isolation of the transconjugants from macerate samples was reduced in some cases. To help elucidate factors that may account for the observed variability of the various transconjugants in bioassays, in the next section these transconjugants and their wildtype counterparts were assessed for their effects on the production of proteases and lipases as outlined in Chapter 2, section 2.3.2.

7.2.5 Visual cell differences of transconjugants

After the construction of the various bacterial transconjugants used in this study, it was observed that two specific conjugants, 477 + pAfpX, and 3041 + pAfpX, exhibited a reduction in colony size on LB agar plates relative to 477 with other plasmids (Figure 75 and Figure 76). This effect only occurred with the addition of the pAfpX plasmid type and only in these two wildtype plasmid-free isolates. To assess the differences and to see whether these posed any large differences in morphology microscopy was performed as described in Chapter 2 section 2.3.7.

There is the potential that these smaller colonies represent a reduced ability for these transconjugants to establish. It was also observed that the time required at 30^{o}C for colonies growth was extended for these isolates, requiring more than one day incubations. As no differences were determined between isolate chromosome (identical wildtypes and pAfpX transconjugant cell morphology) it was concluded that the addition of pAfpX was not having any cell development implications that would affect the individual cell. However, further studies are required to establish why colony morphology differs between the recipient strain.

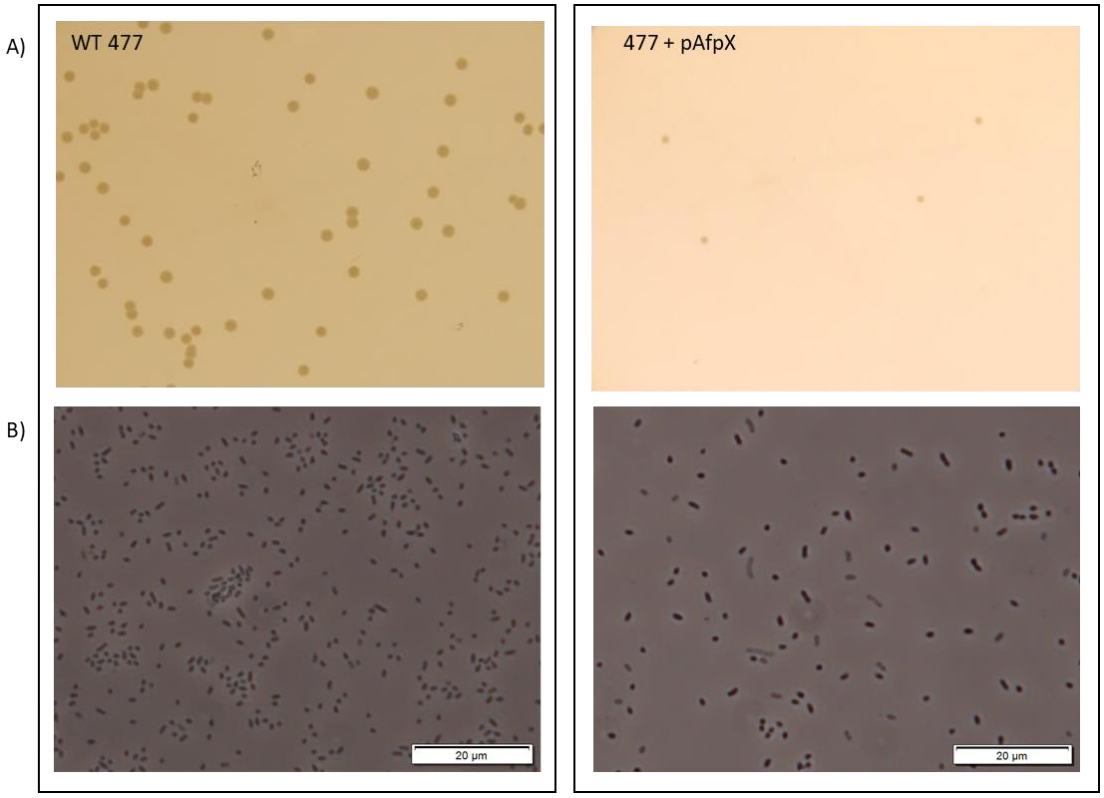


Figure 75 Plate morphology and microscopy of 477 and transconjugant 477 + pAfpX after 24 h growth
Panel A) colonies on LB agar B) Phase contrast microscopy at 100× magnification.

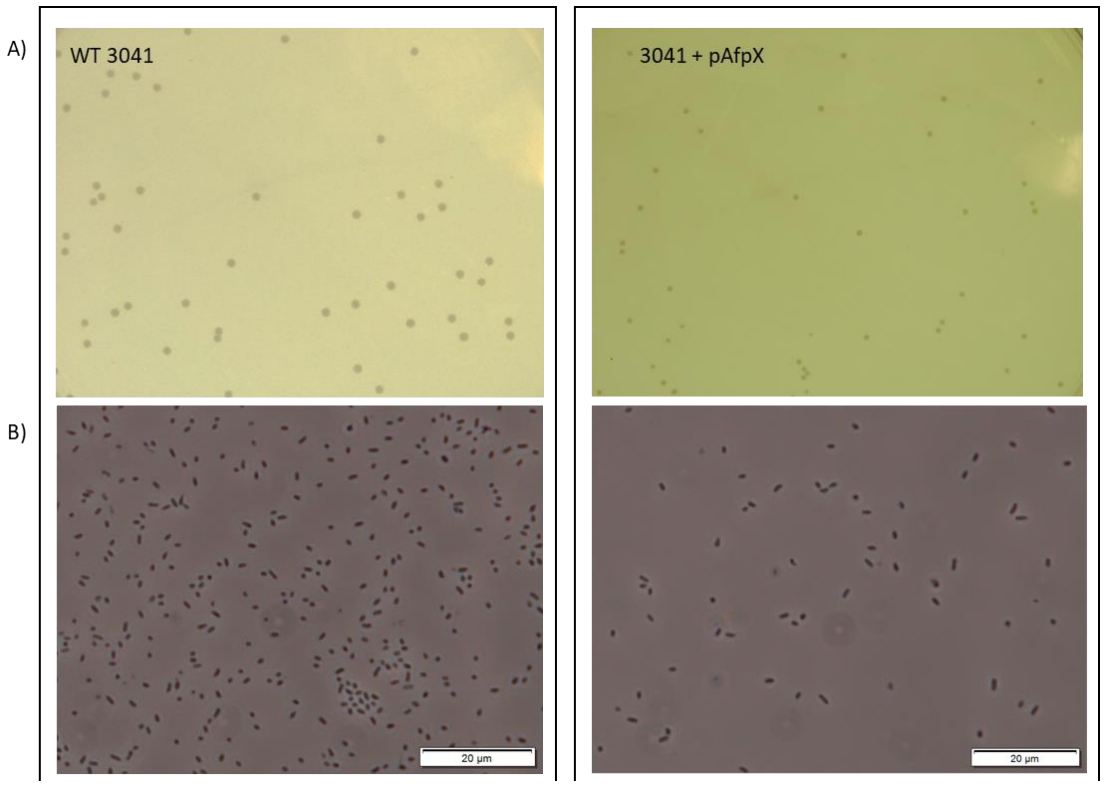


Figure 76 Plate morphology and microscopy of 3041 and transconjugant 3041 + pAfpX after 24 h growth
Panel A) colonies on LB agar B) Phase contrast microscopy at 100× magnification.

7.2.6 Assessment of transconjugant growth relative to wildtype

To determine any metabolic burden potential of plasmids, impose on the non-native chromosomes, growth curves were undertaken in both LB and M9 minimal salts (glucose) media without antibiotics. Overnight cultures were grown as described in Chapter 2, section 2.3.1, with $OD_{600,}$ then checked for each isolate to minimise the difference in starting inoculation (to approximately 0.3-0.4 when diluted 1:10). Dilutions derived from undiluted sample to 10^{-10} dilution, with the final plate well being used as a blank control of the medium analysed. The growth curve was undertaken as described in Chapter 2, section 2.3.6. With reference to Figure 77 (Ai and Bi) and Figure 78 (Ai and Bi), under ideal conditions in LB broth medium, a minimal difference in growth was observed between the wildtype and each transconjugant. The conjugates took over 1 h longer in the lag phase before the exponential growth phase. Trend changes can be seen more dramatically in *S. proteamaculans* 3041 and 3041 + pAfpX, where the addition of pAfpX impeded growth, and in *S. entomophila* iDIA + pAfpX, where peak OD_{600} is lower than for other transconjugants and the WT iDIA (Figure 78B).

The trends observed in minimal stress media i.e., in M9 medium, were more obvious in comparison to LB broth (Figure 77 Aii and Bii, Figure 78 Aii and Bii). iDIA had a notably slower initial exponential growth phase even compared to iDIA + p626, although both these peaked after 24 h at similar OD's. The addition of a plasmid to the *S. entomophila* 477 plasmid-free cell had the opposite effect under stress conditions than seen for *S. proteamaculans* 3041. Initial growth of the plasmid transconjugants (+ p626, pA1M02, or pAfpX) was observed to reach the exponential phase approximately 5 h faster than in the wildtype *S. entomophila* 477 isolate. By 36 h, the wildtype had overtaken the pAfpX transconjugant in relative OD_{600} but was still lagging 477 transconjugants for p626 and pA1M02.

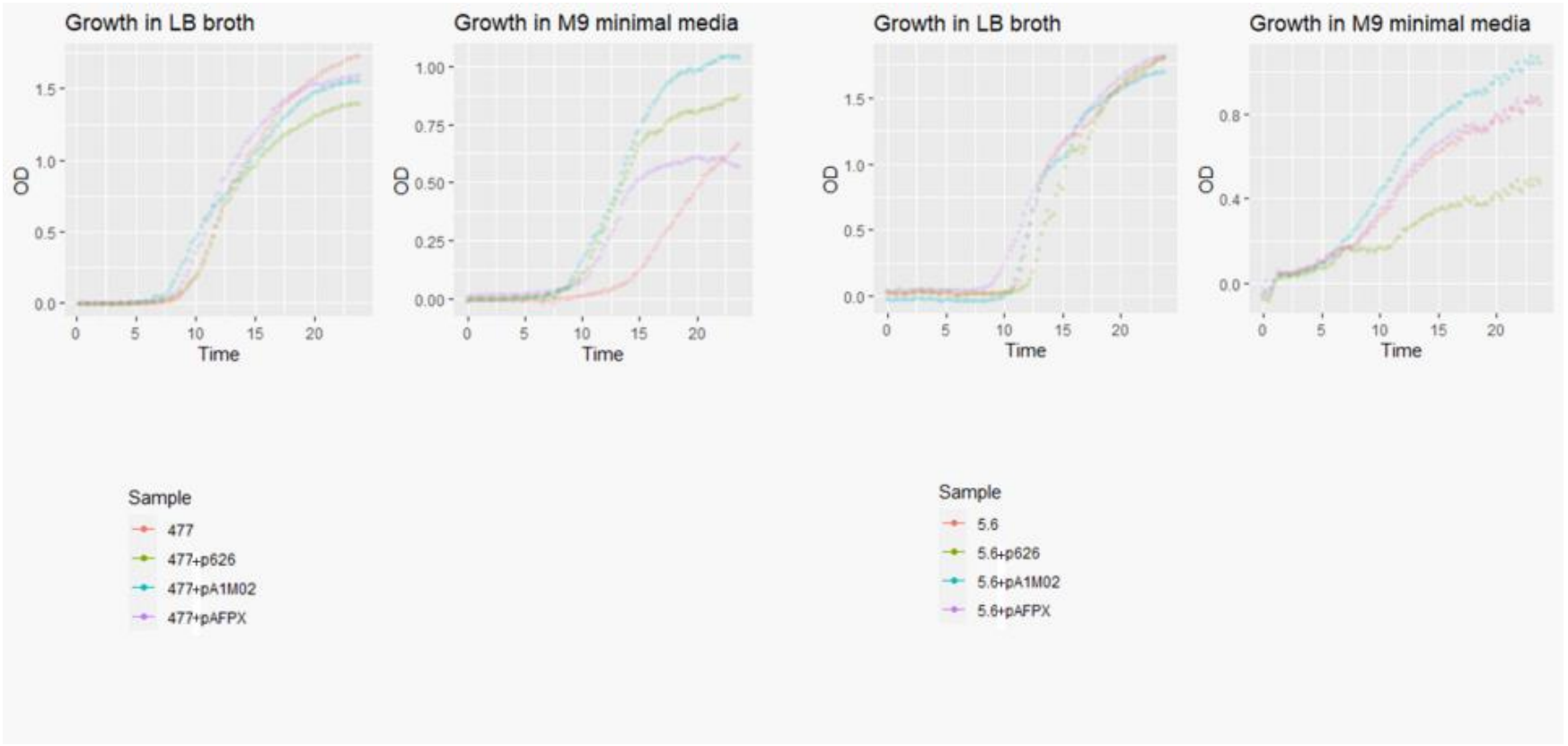


Figure 77 Growth curve of transconjugants and their wildtype counterparts

A) 477 wildtype and its plasmid conjugants in i) LB broth and ii) M9 minimal salts media; B) 477 wildtype and its plasmid conjugants in i) LB broth and ii) M9 minimal salts media. Growth curve conducted over 24 h with measurements take at 15 min intervals.

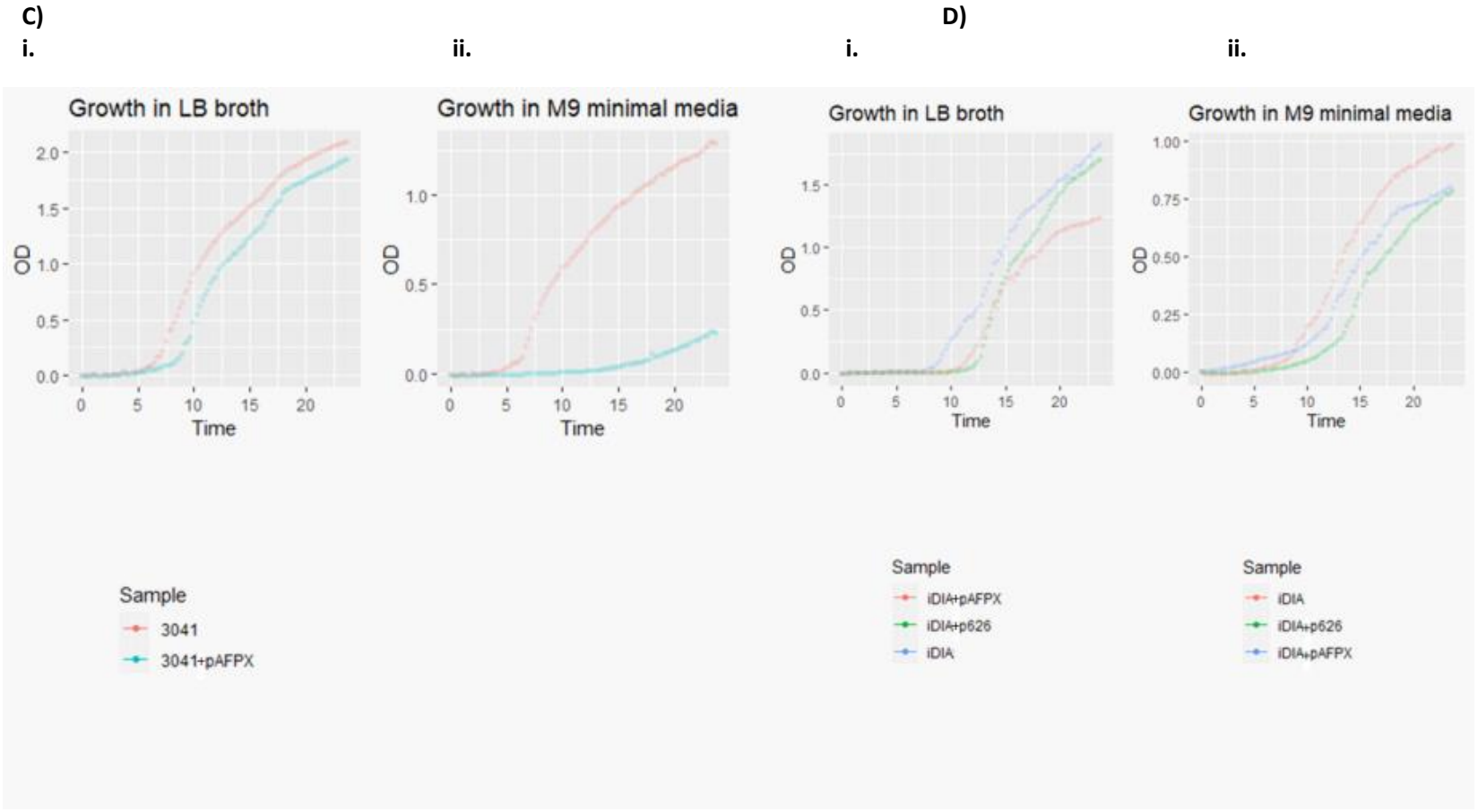


Figure 78 Growth curves of transconjugants and their wildtype counterparts

C) 3041 wildtype and its plasmid conjugant in i) LB broth and ii) M9 minimal salts media; D) iDIA wildtype and its plasmid conjugants in i) LB broth and ii) M9 minimal salts media. Growth curve conducted over 24 h with measurements take at 15 min intervals.

Assessing the plasmid transconjugants of 5.6 (+ pA1M02, pAfpX, and p626 respectively) the showed 5.6 with its wildtype plasmid restored (5.6 + pADAP) outperformed all other transconjugants and the heat-cured isogenic 5.6. This was expected due to the restoration of the wildtype phenotype (Figure 77B). The addition of both plasmids from *S. proteamaculans* AGR96X and *S. entomophila* isolate 626 to *S. entomophila* isolate 5.6 reduced the overall fitness of the 5.6 transconjugants under M9 minimal medium, but little change is observed in nutrient-rich LB broth.

7.2.7 Accessory determinants and plant beneficial traits

By undertaking plate assays (described in chapter 2, section 2.3.2) assessing for the production of secondary accessory determinants the effect that the acquisition of pADAP- like plasmids has on the bacteria's ability to produce various beneficial proteins useful for the potential invasion and colonisation of a host organism or occupation of a specific niche was assessed. In Chapter 4, the use of plate assays revealed that AGR96X has a different profile than the other tested isolates for the production of accessory determinants (Figure 79). Plate assays revealed differences in secretion between species for siderophore and DNase degradation. By defining any differences imparted by a different plasmid on different chromosomal backgrounds, we can postulate whether the observed effects relate to altered virulence or growth of the transconjugants. If the chromosome can influence the expression of plasmid-encoded genes acting as a potential metabolic burden, this would support the observation that pAfpX transconjugants showed a reduced growth rate under stress in comparison to those of pA1M02 and p626 when (section 7.2.6).

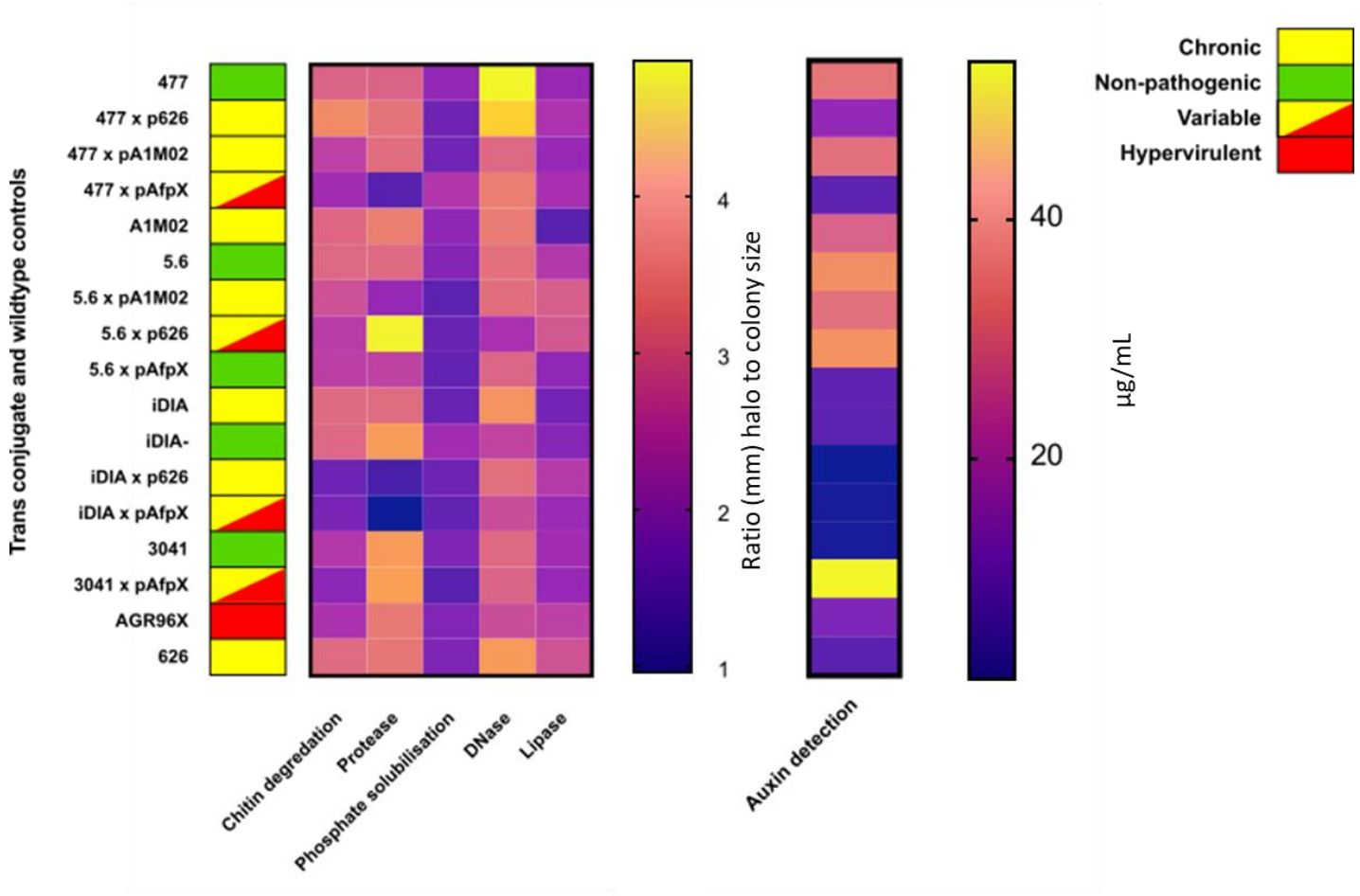


Figure 79 Response of all conjugates and their wildtype chromosomal counterparts to accessory virulence detection and plant growth promotion assays. Plate assay response is defined as the ratio of halo (mm) to colony size (mm). Auxin detection using calibration curve expressed as µg/mL of IAA detected.

To determine if there were any statistical differences in the observed production of accessory virulence factors, a general linear model was constructed between all isolates with plate assay as the variable response. Means were assessed for any difference in significance between the wildtype and their transconjugants. The addition of various plasmid types into plasmid-free *S. entomophila* 477 had a variable effect. A significant difference (P=0.021) in protease production was only observed between pAfpX carrying conjugant (477 + pAfpX) (Figure 80). This was again observed when the response to chitinase degradation was assessed (P=0.003). With the exception of *S. entomophila* transconjugants 477 (p626 and pA1M02), no other bacterial transconjugant was noted as being significantly different from the mean of its wildtype. Assessments of DNase production revealed that although, visually, there was a difference between transconjugant isolates (+ pA1M02, p626, and pAfpX) and the WT *S. entomophila* 477 (Figure 80), the only significant difference was observed between WT 477 and *S. proteamaculans* WT AGR96X (P=0.04). It is assumed the addition of *S. entomophila*-derived pADAP has a negligible effect on the production of accessory virulence factors in *S. entomophila* 477, whereas the opposite is noted for the addition of pAfpX. No significant difference was noted between *S. entomophila* WT 477 and plasmid transconjugants for phosphatase solubilisation or lipase detection. When comparing these transconjugants to the WT plasmid donor isolate there were similar trends. The 477 + pADAP transconjugants sourced from either 626 or A1M02 exhibited similar responses in protease and chitinase degradation (Figure 81 and Figure 82).

Using additional statistical parameters (Tukey's range test), differences in the pathotype of the donor, recipient, and transconjugants were assessed. In the case of 477 and its transconjugants, WT 477 as the non-pathogenic isolate had a statistically larger DNase production (P=0.025) than its three transconjugants (Figure 80), which under the Tukey method for 95% confidence group together in the same level of response. No other plate assay showed any significance with the addition of pathotype and comparison with the plasmid donor types. Auxin detection was also not statistically different between pathogenic and non-pathogenic (P=0.779).

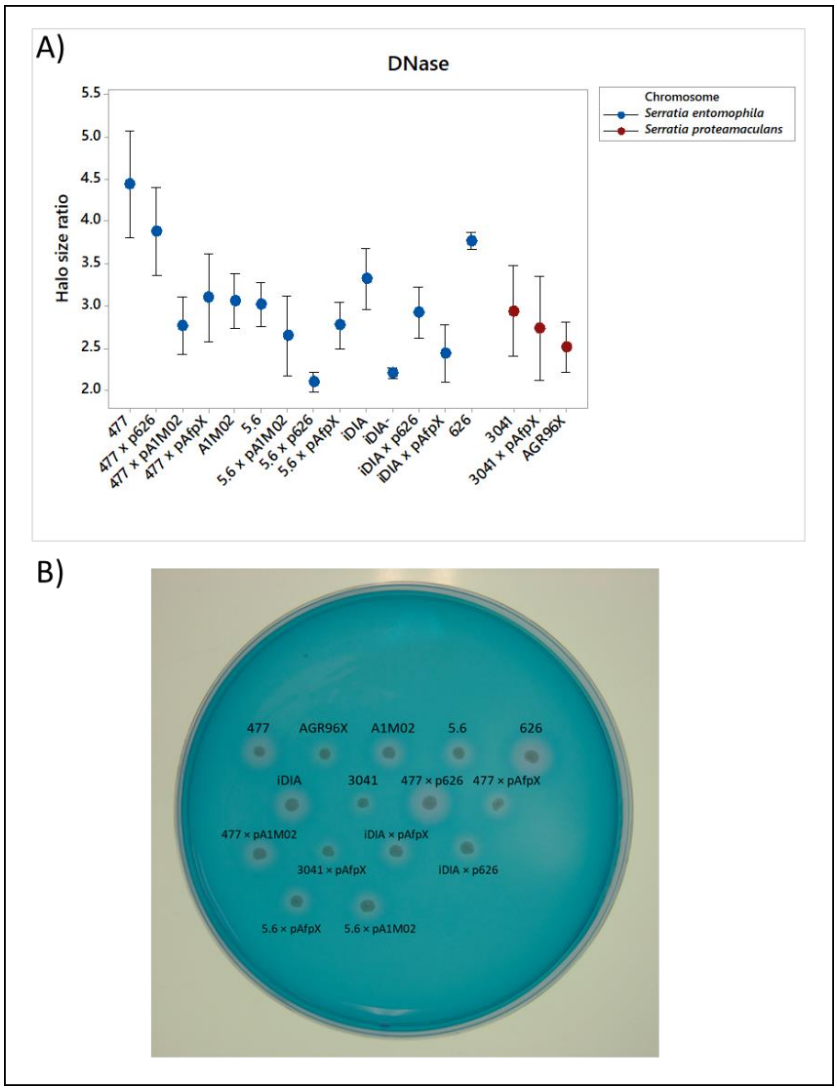


Figure 80 DNase expression across all transconjugant isolates and their wildtype counterparts. A) Average ratio of DNase expression with standard error for all isolates split by species and pathotype. Standard error bars shown. B) Plate showing halo differentiation for isolates assessed.

Isolate 3041 was the only *S. proteamaculans* WT isolate that responded strongly in plate assay comparisons to the conjugation of pAfpX, with similar phenotypes as observed for the AGR96X donor isolate. For the most part, relative to WT 3041, 3041 + pAfpX was relatively stable in its response to the detection of accessory virulence assays. These isolates are both members of the subspecies *S. proteamaculans* subsp. *quinovora*, therefore an association between similar species or subspecies could confer higher stability when conjugated. No statistical difference was observed in protease, lipase, DNase, chitinase, or phosphate solubilisation from either the wildtype plasmid negative isolate or *S. proteamaculans* AGR96X. Of note, an *S. proteamaculans* 3041 + pADAP transconjugant could not be acquired from either the A1M02 or 626 *S. entomophila* donor limiting the 3041 transconjugant comparison to AGR96X. grouped ANOVA comparative analysis was undertaken for pathotype with the addition of AGR96X as the plasmid donor control, grouping AGR96X and 3041 + pAfpX against WT

3041. No significant difference was found for any accessory virulence determinant production using this analysis.

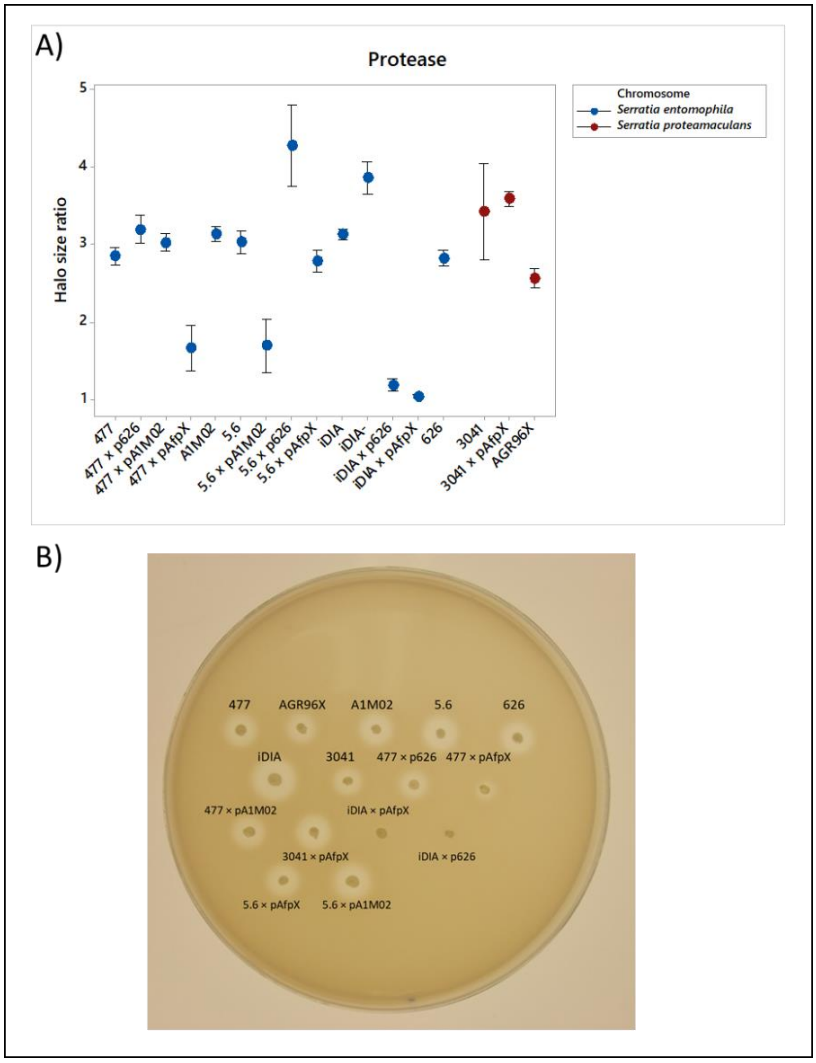


Figure 81 Protease expression across all transconjugant isolates and their wildtype counterparts. A) Average ratio of protease expression with standard error for all isolates split by species and pathotype. Standard error bars shown B) Plate showing halo differentiation for isolates assessed.

In the case of iDIA, similar results were observed to the 477 + pAfpX transconjugant, exhibiting some significant differences to its wildtype AGR96X plasmid donor and wildtype iDIA. There was a variance between both means in protease response (P=0.042) (Figure 81). Both iDIA transconjugants (+ pAfpX and p626) showed a significant change from the wildtype in chitin degradation with significantly lower means (Figure 82). This result demonstrates that the acquisition of a new plasmid can impact the chromosome as having a reduced capacity to release extracellular chitinases (P=0.007 in both cases). No significant differences were observed for phosphate solubilisation, DNase or lipase assays, or related to pathotypes.

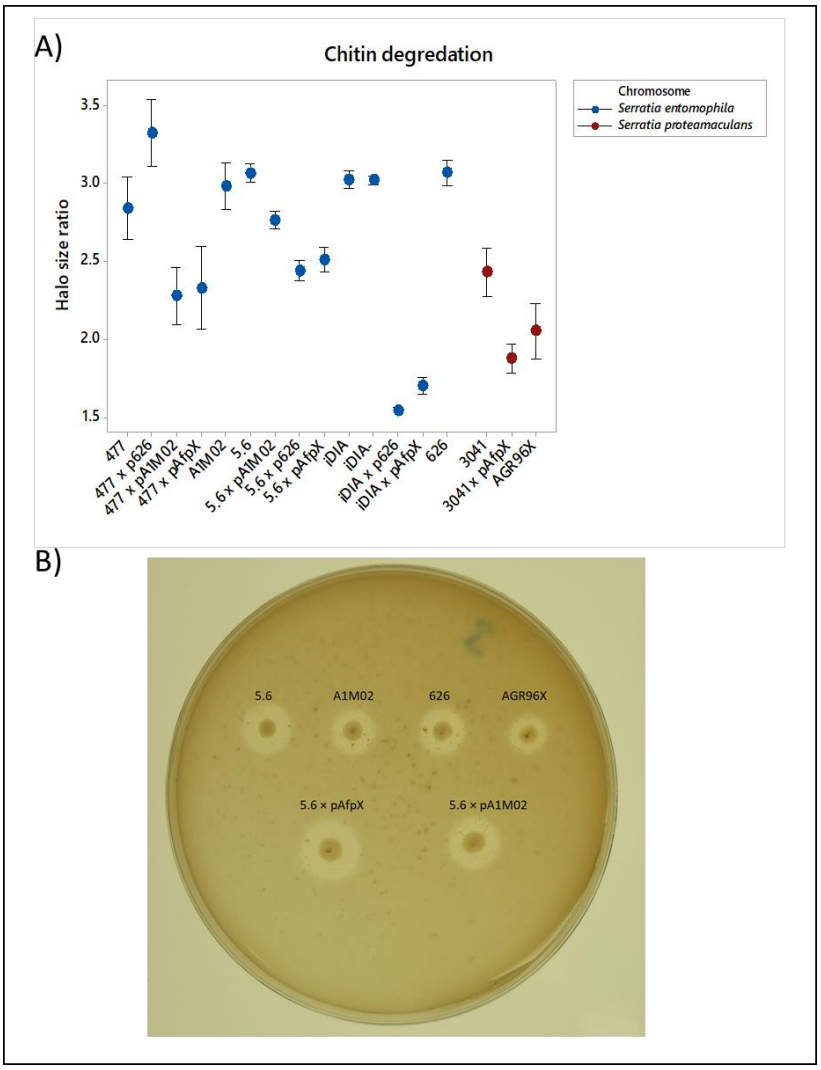


Figure 82 Chitinase degradation assays across all transconjugant isolates and their wildtype counterparts.
A) Average ratio of chitinase expression with standard error for all isolates split by species and pathotype. Standard error bars shown B) Plate showing halo differentiation for isolates assessed.

The controls *S. entomophila* 5.6 and A1M02 were used for the assessment of 5.6 transconjugants. Initial 5.6 + pADAP (plasmid restored) were compared against A1MO2 wildtype and its isogenic heat-cured control, 5.6. Including the further plasmid transconjugants (5.6 + pAfpX and 5.6 + p626), there was no difference observed for chitin degradation and phosphate degradation, or protease and DNase production(Figure 80- Figure 84). Overall, of all transconjugants assessed, those utilising 5.6 (as a group) showed the least amount of change in comparison to their wildtype controls. Transconjugants of 5.6 and the wildtype donor controls were contrasted based on pathotype to show any differences. No significant difference could be found in this manner for the addition of p626 or pAfpX. The restored plasmid transconjugant (5.6 + pADAP) showed no significant differences in the production of chitinases, proteases, phosphatases, or DNases, as expected (Figure 80- Figure 84). As expected, this

indicates that the removal of, and subsequent re-conjugation with, the plasmid does not impact the production of these accessory virulence factors.

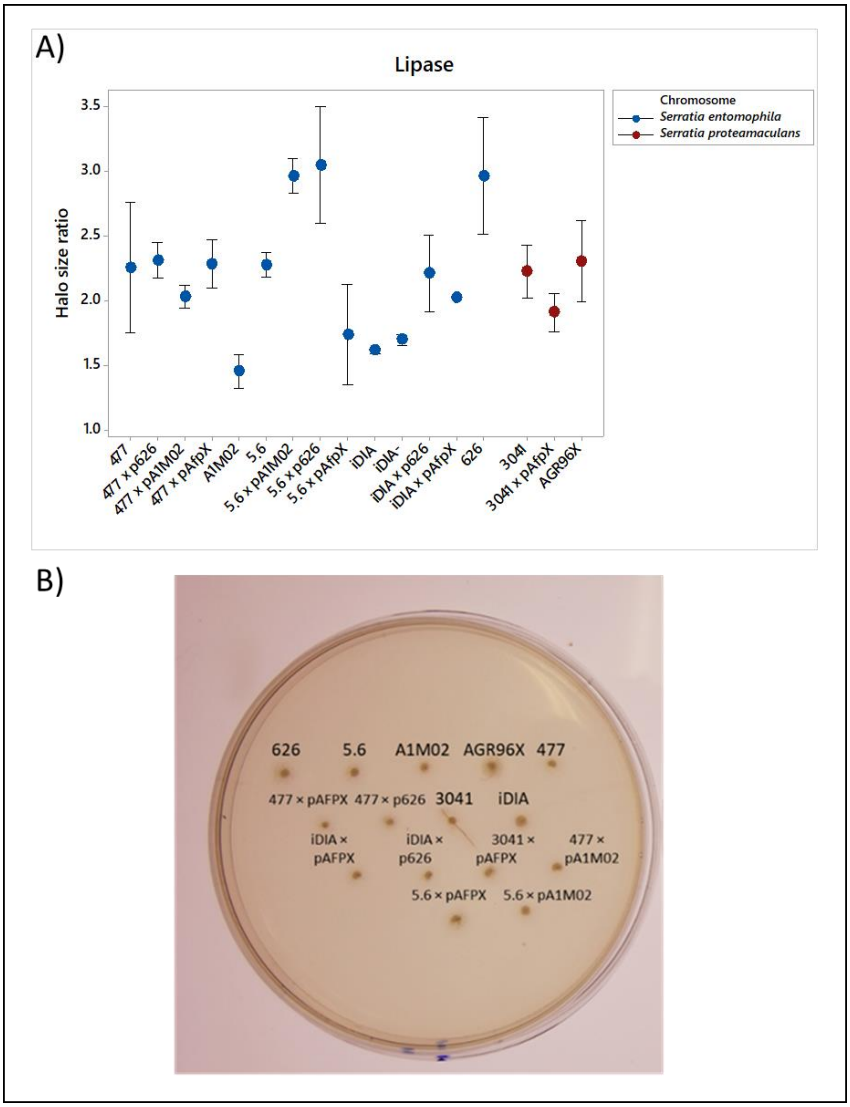


Figure 83 Lipase degradation assays across all transconjugant isolates and their wildtype counterparts.
A) Average ratio of lipase expression with standard error for all isolates split by species and pathotype. Standard error bars shown B) Plate showing halo differentiation for isolates assessed.

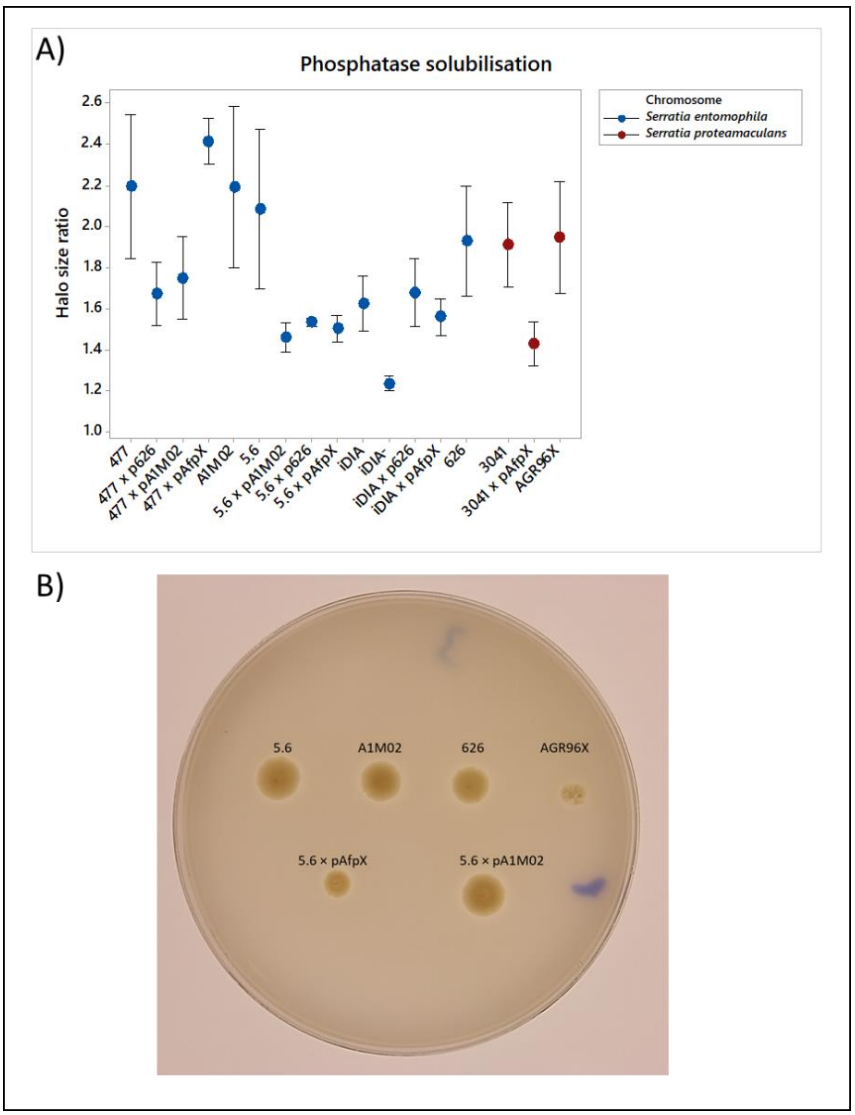


Figure 84 Phosphatase solubilisation assays across all transconjugant isolates and their wildtypes A) Average ratio of phosphatase expression with standard error for all isolates split by species and pathotype B) Plate showing halo differentiation for isolates assessed.

The results from the auxin assays resulted in a no-linear model (Figure 85), and therefore a single hypothesis test generating a P-value could not be used. However, with reference to Figure 88, it is evident that the addition of a non-resident plasmid type does not benefit the production of auxin. The addition of pAfpX reduced auxin production encoded by all *S. entomophila* chromosomes but interestingly seems to increase overall synthesis in *S. proteamaculans* 3041. The addition of variant pADAP plasmids to *S. entomophila* isolates iDIA and 477 also reduced the ability to produce auxin, although this is less defined than with the addition of pAfpX.

To encompass the whole dataset as one largescale comparison, pathotypes used for individual analysis were then compiled to generate a large linear model before individual comparisons using the Tukey

comparison at 95% confidence. Although through the assessment of interval plots of auxin production in transconjugants a trend was observed, no significant link was found between the pathotype of a wildtype isolate or transconjugant and the production of accessory virulence determinants, which corroborates the results reported in Chapter 4 where there was a tenuous link between IAA detection response assay and virulence capacity.

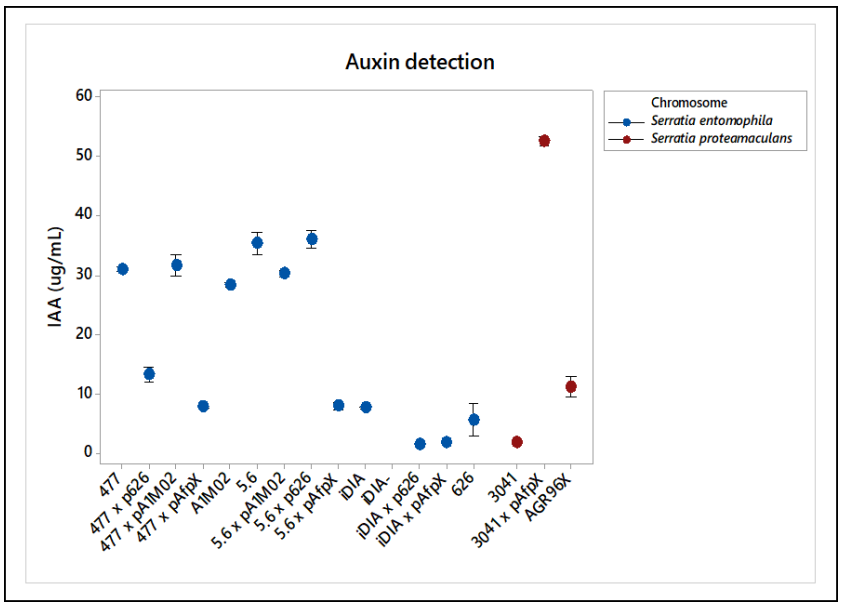


Figure 85 IAA (auxin) detection assays (µg/mL) across all transconjugant isolates and their wildtype counter parts.

7.3 Discussion

Through the assessment of eight isolates of *S. proteamaculans* and *S. entomophila* for their ability for causing disease after induction with MitC (a stress agent inducer of the SOS response), only AGR96X was found to be a consistently highly inducible isolate. Variability in bioassay results was observed over the years and with different collection sites for grass grub. Degrees of susceptibility to *Serratia* spp. inducing amber disease may fluctuate between populations and yearly cycles. Susceptibility to pathogens has been demonstrated to fluctuate seasonally (Dowell 2001). Yearly variations in larval age and potentially genotypic variation (such as with subspecies) may account for some of these changes. The population of grass grub sourced from various locations around New Zealand has yet to be fully characterised, so the extent of subspecies is yet unknown. Initial studies by Richards et al. (1997) showed that diverse genetic variation could be found when amplifying DNA from samples of grass grub from across New Zealand, with distinct differences in the North and South islands. More recently, Lefort et al. (2013) used molecular-based methods to differentiate between early instar larvae of *C. giveni* and *Costelytra brunneum*, another endemic scarab that bears a striking morphological similarity.

Further characterisation would therefore expand on this knowledge and benefit the establishment of microbial interactions in grass grub.

The SOS response is regulated on the bacterial chromosome by *recA*-*lexA*, where *lexA* in uninduced cells represses genes based on the presence of an SOS BOX operator system binding upstream of the specific gene or operon. DNA damage either from natural means or MitC induction facilitates *recA* into activation that then auto digests *lexA* ability to bind and repress these genes (Sutton et al. 2000). In the *S. proteamaculans* hypervirulent isolate AGR96X, induction of the SOS response has been shown to facilitate the overproduction of AfpX particles (Hurst et al. 2018). In the case of other isolates, the induction of virulence factors is unknown. Further analysis including purification and visualisation of these virulence factors with transmission electron microscopy (TEM) could then identify these induced factors as Afp or non-Afp related. Until then, this induction cannot be limited to only AfpX and may extend out to TC cluster or Sep complex. Further investigation would be required to fully define the mechanism at play, but it is observable that no other isolate has the same consistent response as *S. proteamaculans* hypervirulent isolate AGR96X.

When these experiments were designed and implemented, the isolates utilised were split evenly between *S. proteamaculans* and *S. entomophila*. As later bioinformatic analysis then reclassified some isolates, the sample size became more skewed to *S. entomophila* isolates. Based on the findings presented in section 7.2.1, it would be useful therefore to undertake more experiments using additional isolates of *S. proteamaculans* to define if, similar to AGR96X, other isolates are also MitC inducible. If isolates were found to be inducible, cross-referencing the putative SOS BOX predictions between AGR96X and other inducible isolate sequences could elaborate on the potential gene mechanism regulating the induction of AfpX from the chromosome Incorporating RNA transcriptome analysis of induced and noninduced isolates would also serve as a good comparative base for determining these mechanisms. As discussed earlier, current work by Hurst (unpublished data) seeks to determine these mechanisms by utilising transcriptome data of MitC induced and non-induced isolates of *S. proteamaculans* and *S. entomophila*. This data should then complement what was found in this study.

Several SOS BOX predicted regions contain genes that regulate transporter proteins and phage packaging genes, though none of these seems to have a role in the regulation of AfpX production. However, through *in silico* assessments, several SOS BOX motifs were predicted on transcriptional regulators which in turn could indirectly affect AfpX or other virulence factor production. In *Shigella flexineri*, the prophage-encoded gene encoding a glucosyltransferase located on the chromosome was implicated in the regulation of pathogenicity by the type II secretion system (TIISS) located on the

virulence plasmid. In this instance, losing glycosylation masked the TIISS by forcing the O-antigen to protrude from the bacterial cell, resulting in a reduction of pathogenicity (Nhieu and Sansonetti 1999). This example demonstrates how secondary mechanisms can influence virulence that is co-located on the plasmid, potentially explaining how no specific virulence-associated gene was associated with AGR96X induction.

The action of autodigestion of *lexA* by *recA* has been previously implicated in actively switching the growth cycle of λ phage in *E. coli* from lysogenic to lytic growth (Little 1984). In *S. proteamaculans*, the activation of AfpX production is mediated by the chromosome, so any repressor action would have to be *in trans* with the plasmid. Plasmid regulation by the chromosome has been described and is relatively well known in other systems. In *Agrobacterium tumefaciens*, the chromosomally located *chvL/chvG* sensor regulator system is essential to the expression of virulence (Charles and Nester 1993). Potentially, one of the regulators found encoding a downstream putative SOS BOX could then regulate phage production by the AfpX or initiated the lytic cycle of the phage, resulting in an overproduction of AfpX resulting in the high response seen that isn't observed in other isolates.

As previously found by Grkovic et al. (1995), plasmid conjugation confers the ability to cause amber disease in grass grub larvae. We see that although this pathogenicity is conferred in transconjugants, the limitations in comparison to the wildtype pathogenic isolates are confined to an erratic pathotype (with variable disease induction). Some chromosomal backgrounds do not confer the same level of metabolic fitness to the plasmid carrying wildtype. When compared to wildtype AGR96X, conjugation of pAfpX into *S. entomophila* chromosomal backgrounds reduced overall virulence. This reduced virulence phenotype implicates the chromosome in contribution to the disease pathotype. These findings corroborate what was found by Sitter (2020) and Grkovic et al. (1995), where pathogenicity was conferred in plasmid transconjugants but to varying degrees against the wildtype. These results implicate an as yet unidentified chromosomal mechanism that imparts full pathogenicity or a reduction in the ability for the bacteria to effectively adhere to the larval hindgut and establish a successful infection by the reduction in the production of proteases and fimbriae. Relating these results to what was found *in silico* in Chapters 5 and 6, the differences observed on the chromosome is small and found no such gene(s) associated with pathotype. There is potential that fitness benefits in *S. proteamaculans* to the extra carbohydrate metabolism pathway (Doudoroff pathway discussed in Chapter 6) could be beneficial in environments lacking the requisites for potential replacement itaconate degradation pathway in *S. entomophila*. The addition of pAfpX from hypervirulent AGR96X did not confer the same levels of pathogenicity into any of its variant transconjugants. Bioassay results in this study for the LC_{50} of AGR96X were a log-fold higher than those found by Hurst et al. (2018) at $4.89 \times 10^3 \pm 0.92 \times 10^3$ cells, although LT_{50} results were highly similar (3 days). This trend was followed with the *S. entomophila*

isolates, where previous LC_{50} assays showed 2.95×10^4 cells (Jackson et al. 2001), which was only observed in iDIA. The addition of pAfpX conferred less consistent results than that of standard pADAP from *S. entomophila* isolates 626 and A1M02, implying there is a degree of chromosomal selectivity not seen in the promiscuous broad-host range AGR96X in plasmid acceptance. The addition of pAfpX resulted in an increased time for 50% of grub to become diseased (LT_{50}) in all transconjugants, with isolates iDIA and 3041 taking three days longer compared to *S. entomophila* 477. Again, the introduction of AfpX resulted in primarily a two log-fold increase of overall lethal concentration except in iDIA, where a one log-fold increase was observed. Overall, it is clear that pAfpX has a further burden on the chromosome increasing LC_{50} as well as reducing the production of chromosomally encoded accessory virulence determinants compared to the addition of pADAP.

By assessing the expression of degradative enzymes such as DNases, proteases, and lipases in transconjugants the impact the plasmid has on the bacterial ability to exhibit natural functions. The addition of the plasmid from AGR96X confers a degree of reduced accessory enzyme production especially in *S. entomophila* 477 + pAfpX. Previous studies utilising phenotypic assays in unison with metabolomic analysis have found disjointed results facilitating the conclusion that the fitness burden and benefits of plasmids are multifactorial (San Millan et al. 2018). This is again supported by a study by Anthony Mason and Bailey (1989) which found that plasmid presence in plasmid-bearing isolates affected the activity of certain enzymes when compared with that of plasmid-free isolates in *E. coli*. Here we see that, particularly with the addition of pAfpX, a reduction in the production of these accessory determinants was observed revealing the degree of burden the plasmid imparts on the cell. With the addition of pADAP from 626 and A1M02, the results were less clear cut and this likely reflects the high nucleotide identity of p626 with pADAP. The reduction in the synthesis of these accessory determinants could be detrimental to the infection process, causing the variability seen in the prior mentioned bioassays conducted. The production of lipolytic enzymes can facilitate the invasion of the host by a pathogenic cell by destroying host cell tissue and evading the host immune response. Reduction of this production would generally lead to a reduction in in-host survival. Further analysis into the regulatory mechanism that mediates these changes will be reported in Chapter 8.

The metabolic burden of the addition of plasmids in plasmid-free backgrounds has been shown in this study. Previous work found in *Pseudomonas aeruginosa*, the burden of plasmids was multifactorial (San Millan et al. 2018). The addition of 'domestic' plasmids that had similar G+C and codon use profiles to the chromosome, added few molecular alterations, and resultingly expressed low levels of plasmid-related genes, has the least fitness impacts. This corroborated the results found in this study, where chromosomal backgrounds of *S. entomophila* (5.6, iDIA, 477) acquired *S. entomophila* pADAP plasmid types with ease and a less obvious burden in growth, more so than the AfpX plasmid The *S.*

entomophila isolate MC2, sourced from the grass grub-free region of the West Coast of the South Island, proved difficult to conjugate. This finding may reflect the genomic similarities between MC2, and Chatham Island isolate 440 outlined in Chapters 4 and 5. This implies a potentially shared aversion to the acquisition of the pADAP plasmid by these non-pADAP plasmid-bearing isolates. This could also be a representation of the overall low rates of conjugation. The addition of pAfpX to *S. proteamaculans* wildtype plasmid-free isolate 3041 produced conflicting results, where the addition of the plasmid also had negative effects when grown under stress. The addition of the plasmid to the wildtype plasmid-free background (3041 + pAfpX) overall may have a negative effect which is entirely separate from the G+C content and codon-use advantage for similarly constructed chromosomes.

Similar isolates such as *S. entomophila* isolates 626 and 477 seem to have different mechanisms for conferring sustainable relationships with plasmids. Isolate 626 as previously discussed has a naturally occurring pADAP and isolate 477 is naturally plasmid-free. Although isolate 477 is a naturally occurring plasmid-free isolate, it was found that the cell readily took the pADAP plasmid via conjugation, showing the compatibility with the chromosome. A 'source-sink' relationship between bacterial communities and plasmid acquisition has been previously described in a study by Hall et al. (2016), where long-term stability of a plasmid in a population was maintained by *P. fluorescens,* but acquired and subsequently lost in *Pseudomonas putida* populations. The potential for a similar relationship in isolates used in this study is there. Chromosomally, the major differences between these two isolates were outlined in Chapter 5, showing little difference asides from two 40 Kb regions of variation. Some isolates were not able to accept plasmids under experimental conditions. The inability to conjugate pADAP into AGR96X mutant XΔNOVA supports the handcuffing incompatibility mechanism discussed by Sitter (2020), where the pADAP from *S. entomophila* was not able to replace the pAfpX pADAP variant. Plasmid incompatibility groups are well described in the Enterobacteriaceae, with 27 groups currently defined (Carattoli 2009). As these incompatibility systems are mediated by the plasmids themselves, no differences could be found on the chromosome of isolates encoding plasmids unable to accept pADAP. In addition to not accepting based on forced conjugation, some isolates could not be plasmid cured (*S. proteamaculans* AGR96X, *S. entomophila* 440 and 626). Multiple methods of plasmid curing were attempted but all failed to remove the resident plasmid. Some plasmids can encode for plasmid addiction systems (PAS) that act as post-segregational killing elements, killing cells that have lost plasmids. Sitter (2020) identified several novel putative exotoxin effectors present on some pADAP plasmid types, theorised as acting as bacteriostatic TA systems, which instead of inducing death facilitate growth arrest. These systems could potentially be the mechanism impeding plasmid loss in isolates of relevance such as the hypervirulent *S. proteamaculans* AGR96X. Chatham Island isolate 440 is more puzzling in its inability to be conjugated. The large plasmid located in the cell of 440 is not a

pADAP like plasmid, therefore this potentially cannot be attributed to handcuffing, although characterisation of the *recA* gene has not been established.

Variable long-term plasmid persistence within populations of the same species has been previously described in promiscuous plasmids of *Pseudomonas* (De Gelder et al. 2007). De Gelder et al. demonstrated that over 80 generations rapid plasmid loss was observed in some isolates whereas some showed continuous stability over 200 generations, which was attributed to high plasmid cost, or frequency of plasmid segregation. High plasmid cost would be supported by the results in the current study. Plasmid acquisition in some isolates was shown to result in a drastic loss in the capacity to produce accessory virulence determinants, thereby potentially reducing the efficacy of infection as well as impacting cell metabolism. This may begin to explain why we see in *S. entomophila* and *S. proteamaculans*, where wildtype plasmid-bearing and plasmid-free isolates co-exist in the same niche. This leads to the conclusion that a regulatory mechanism confers advantageous retention of plasmids in some isolates but not others. Isolate specific chromosomal traits should therefore be investigated further. Transcriptome data will be presented and discussed concerning this hypothesis in Chapter 8.

Chapter 8

Transcriptional regulation of non-native plasmid transconjugants

8.1 Introduction

The ability of *S. entomophila* and *S. proteamaculans* to transfer pathogenicity by the conjugation of a pADAP family plasmid has been previously assessed and shown by Grkovic et al. (1995) and Sitter (2020). Sitter, however, noted that the conjugation of pADAP (pAfpX) plasmids often resulted in inconsistent disease pathotypes where the transconjugant was less virulent than the plasmid donor wildtype. This was also found in this study (Chapter 7). *In vitro* enzyme plate assays and *in vivo* bioassays described in Chapter 7, revealed differences in the production of accessory virulence determinants when pADAP plasmids were conjugated between differential chromosomal backgrounds. The conclusions derived from these assays were that an as yet unknown chromosomal regulatory mechanism alters gene expression. This may then favour the retention of plasmids in some chromosomal backgrounds and not others. This data further alludes to the coevolution of plasmids to a specific chromosome.

Unsurprisingly, a large proportion of gene expression burden is through the depletion of transcriptional resources, with the introduction of extra expression material in mutated constructs resulting in global regulatory changes (Qian et al. 2017). Plasmids of *Pseudomonas aeruginosa* were shown by San Millan et al. (2018) to alter the expression of common metabolic pathways through burden-driven feedback. This was explored by Ceroni et al. (2017) using *Escherichia coli* models, to investigate the transcriptomic burden of expressing synthetic gene constructs inserted into the genome. Ceroni et al. (2017) found that, regardless of content, expression constructs utilised a high proportion of host cell transcriptional resources. It can therefore be hypothesised, that the addition of a non-native plasmid to a chromosome specialised as a non-pathogenic organism could introduce an excessive burden on host cell polymerases and regulatory machinery that would result in a globally observed effect on expression.

In this Chapter, transcriptome data from selected representative transconjugant isolates and their wildtype counterparts were used to help understand the regulatory burden of the plasmid on non-native chromosomes. To further explore the hypothesis that the plasmids of *S. entomophila* and *S. proteamaculans* have coevolved with bacteria, comparative transcriptomics was undertaken of

selected transconjugants (based on plate assays and bioassays carried out in Chapter 7) 477 + pA1M02 and 477 + pAfpX in comparison to wildtype isolates 477 and A1M02. It was expected that by undertaking these analyses, the metabolic burden on the chromosome could be characterised in relation to the production of accessory virulence determinants, and plasmid-associated insecticidal. Transcriptomic assessments were undertaken in the final stages of the PhD due to delays caused by SARS-CoV-2 lockdowns. Due to time constraints analysis was restricted to the analysis of accessory virulence factors relating to the prior *in vitro* and *in silico* analysis, and burden on pADAP encoded genes.

8.2 Results

8.2.1 Regulation of genes of interest

Based on the hypothesis that a regulatory mechanism is impeding the full expression of virulence in bacterial transconjugants, RNAseq (Chapter 2, section 2.5.17), was undertaken on wildtype isolates 477 and A1M02 and transconjugants 477 + pAfpX and 477 + pA1M02 at time points between 8 and 9 h where OD_{600} showed similar values. Gene expression level was assessed at stationary growth phase cell density of 7.2×10^9 CFU/mL for 477, 6.1×10^9 CFU/mL for 477 + pA1M02, 6.8×10^9 CFU/mL for 477 + pAfpX and 7.5×10^9 CFU/mL for wildtype A1M02. For sample quality control in relation to transcriptome samples, principal component analysis (PCA) was used to show the clustering of replicates in each comparative analysis. The clustering of duplicate samples was confirmed (Figure 86 and Figure 87). Panel B of Figure 86 and Figure 87 shows the number of significant genes of differential expression relative to the wildtype control. The addition of pAfpX has an increased differential expression rate compared to that of the introduction of pA1M02. This would implicate the addition of pAfpX in altering the transcriptome of isolate 477 more heavily, which would be expected when relating to the *in vivo* bioassay and plate assay results reported in Chapter 7.

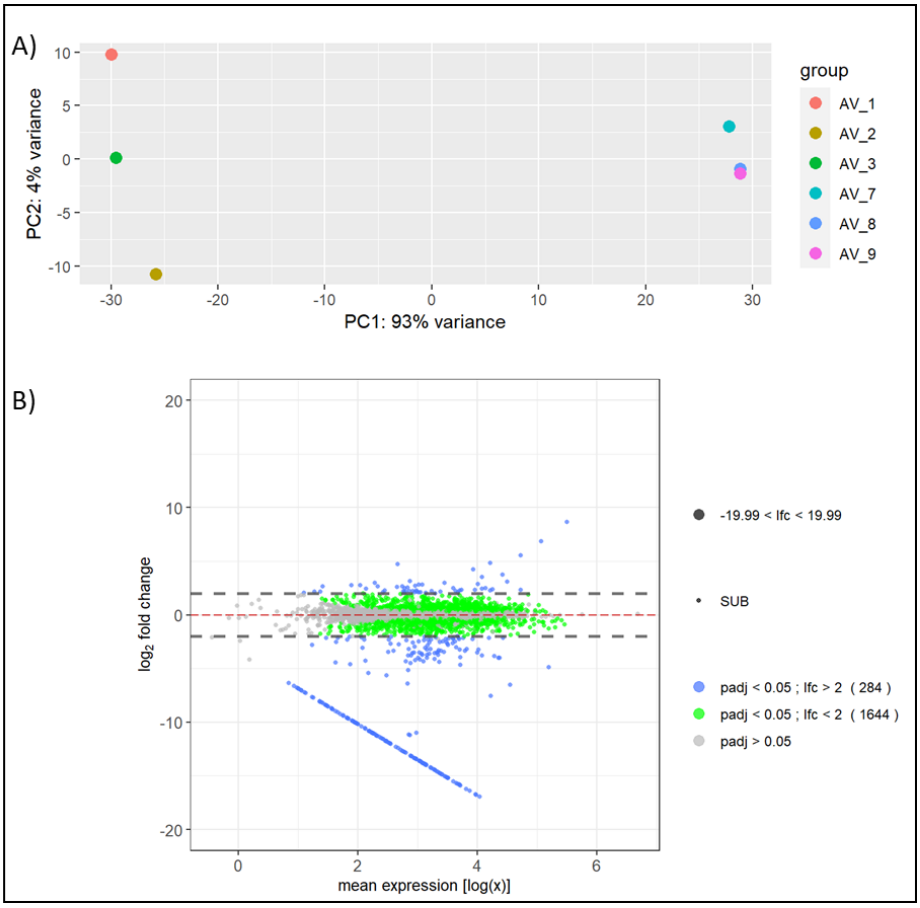


Figure 86 Analysis of transcriptome data in triplicate comprising samples AV1-3 (WT 477) and AV7-9 (477 + pA1M02)

A) Principal component analysis of 477 and 477 + pA1M02 library of normalised data. B) Transcriptome data analysis between 477 and 477 +pA1M02, showing differential gene expression between the transconjugant 477 + pA1M02 and the wildtype 477. Blue dots denote genes with significant changes in expression with an adjusted P value cut off 0.05, log2fold change more than 2. Green dots denote significant changes (P>0.05) of genes with log2fold less than 2. Grey dots show genes with no significance. Downregulated genes in the transconjugant are shown below the red line and upregulated genes in the transconjugant above the red line. The blue line of values represents genes with relatively high expression in one condition compared to the other (where 0, or close values of are found).

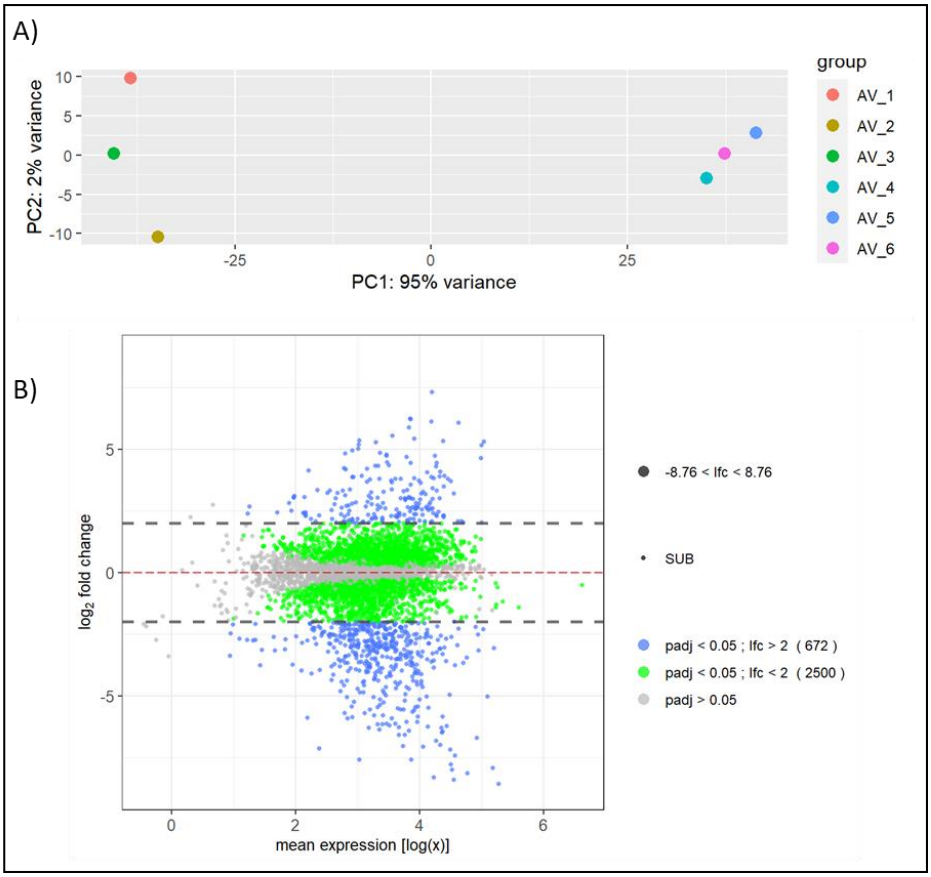


Figure 87 Principal component analysis of in triplicate comprising samples AV1-3 (WT 477) and AV4-6 (477 + pAfpX)
A) Principal component analysis of 477 and 477 + pAfpX library of normalised data. B) Transcriptome data analysis between 477 and 477 +pAfpX, showing differential gene expression between the transconjugant 477 + pA1M02 and the wildtype 477. Blue dots denote genes with significant changes in expression with an adjusted P value cut off 0.05, log2fold change more than 2. Green dots denote significant changes (P>0.05) of genes with log2fold less than 2. Grey dots show genes with no significance. Downregulated genes in the transconjugant are shown below the red line and upregulated genes in the transconjugant above the red line.

Transcriptome data was then assessed to infer the relationship between the addition of a plasmid type to a non-pathogenic chromosomal background and the genes of interest found *in silico* in Chapter 5, corresponding to the plate assay results conducted in Chapters 4 and 7. These analyses focused on the accessory virulence determinants found in wildtype plasmid-free isolate 477 and wildtype pADAP bearing A1M02, to ascertain the burden or benefit the plasmid has on the expression of these enzymes. Comparisons between 477 + pAfpX and AGR96X could not be achieved due to the dissimilarity of the transcript assemblies, therefore 477 + pAfpX was compared to 477 as a baseline.

Assessment of *nucA*, an endonuclease of *Serratia* spp., found that *nucA* was downregulated to log2fold significance >2 with a P-value of <0.05 in both transconjugants (477 + pA1M02 and 477 + pAfpX) relative to wildtype 477 (Figure 88). This indicates the expression of the *nucA* endonuclease is affected by the introduction of a non-native plasmid type. This is corroborated by the relatively low level of expression in the pADAP encoding A1M02, which shows no statistical difference to that of 477 +pA1M02. This links to the plate assays undertaken in Chapter 4, which showed that there was a significant difference between pathogenic and non-pathogenic isolates in the production of DNases. It can be observed here that transconjugants have reduced expression of *nucA*, much like that of wildtype A1M02.

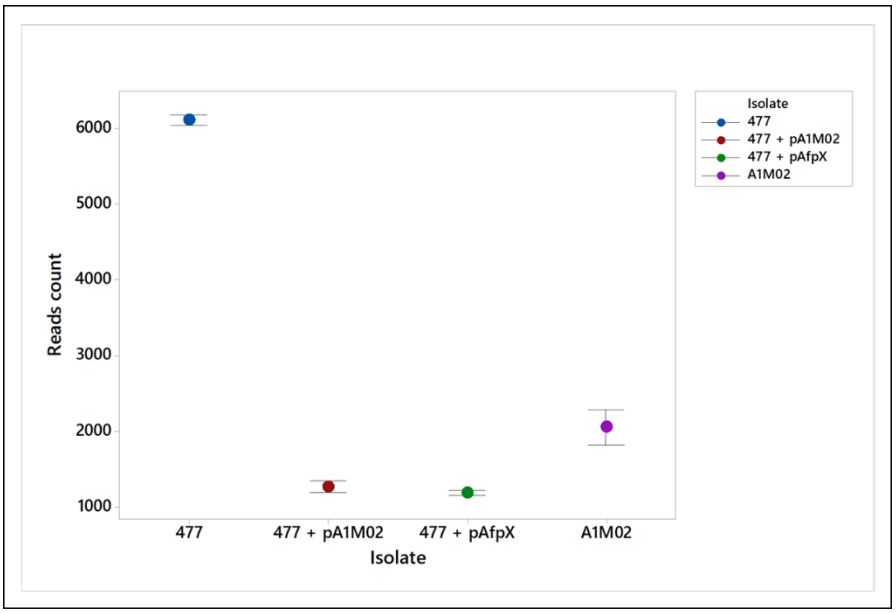


Figure 88 Transcriptome reads count of *nucA* nuclease in transconjugants 477 + pAfpX and 477 + pA1M02 compared to wildtype 477.
Read counts in triplicate. Error bars show standard error of read analysis.

Based on research by Jin et al. (1996), who defined *nucC* as a transcriptional regulator essential for *nucA* expression, transcriptional activity for the homolog of *nucC* in *S. entomophila* 477, and transconjugants 477 + pA1M02 and 477 + pAfpX was assessed (Table 42). *nucC* was found to be downregulated in 477 + pA1M02, but no significant differential expression of *nucC* was found in 477 + pAfpX. The lack of a trend found in 477 + pAfpX implies that this downregulated expression of *nucA* observed in the 477 + pA1M02 transconjugant is independent of the *nucC* transcriptional regulator.

Table 42 Log2fold change observed for regulator, *nucC*, associated with *nucA* expression in relation to wildtype isolate 477

Isolate	LOG2FC	P-value
477 + pA1M02	-0.63	**0.001**
477 +pAfpX	0.015	0.95

Significant log2fold changes underlined (<2)
Significant P-values are shown in bold

Investigations of chromosomal encoded chitinases focused on those identified in isolate 477 in Chapter 5, pertaining to *chiD, chiA, chiA1,* and *chiB* (Figure 89 Transcriptome reads count of chitinase genes in transconjugants 477 + pAfpX and 477 + pA1M02 compared to wildtypes 477 and A1M02. Read counts in triplicate. Error bars show standard error of read analysis.). In both 477 (+pA1M02 and +pAfpX), *chiA, chiB* and *chiD* are downregulated to log2fold >2 relative to wildtype 477. Isolate 477 + pAfpX shows a significant downregulation of the *chiA* gene to log2fold <2 when compared to wildtype 477, but less downregulation than the aforementioned *chiB,* and *chiD* in addition to *chiA1* in relation to wildtype 477. However, in 477 +pA1M02, *chiA1* is upregulated to log2fold <2 relative to the wildtype. No comparison to A1M02 could be undertaken for *chiD*, as wildtype A1M02 lacks this gene. However, comparisons for *chiA, chiA1,* and *chiB* show a log2fold <2 significant change in expression relative to the transconjugant 477 + pA1M02. This implicates the addition of a non-native plasmid to isolate 477 could have a metabolic burden on the chromosome reducing the expression of chitinase genes in comparison to both wildtypes, perhaps as a compensation method to carrying plasmid burden. These findings are in agreement with plate assays conducted in Chapter 7 (section 7.2.7).

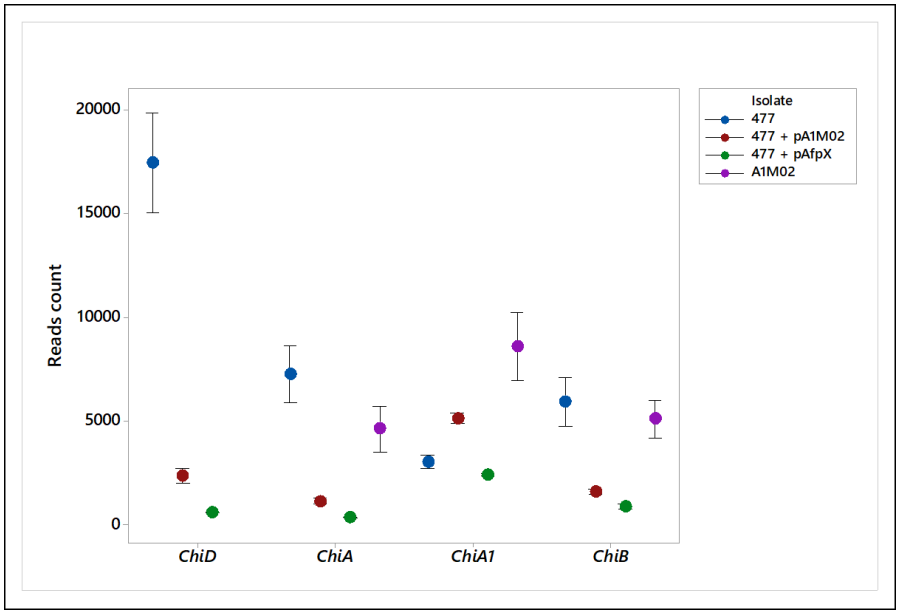


Figure 89 Transcriptome reads count of chitinase genes in transconjugants 477 + pAfpX and 477 + pA1M02 compared to wildtypes 477 and A1M02.
Read counts in triplicate. Error bars show standard error of read analysis.

Two chitin-binding proteins termed chitin binding protein 1 and chitin binding protein 2 (*gbpA_1* and *gbpA_2*) were identified on the chromosome of isolate 477 (Chapter 5). Analysis of the regulation of these genes was conducted on the transcriptome data for the wildtype alongside the plasmid transconjugants (Figure 90). *GbpA_1* showed no significant difference to the wildtype in 477 +pA1M02, whilst log2fold change >2 was observed in 477 + pAfpX indicating significant downregulation. However, for *gbpA_2* both plasmid transconjugants showed as log2fold change >2 of downregulation relative to wildtype 477. Transconjugant 477 + pAfpX showed a log2fold change of downregulation of 6.06 to the 3.2 shown by 477 + pA1M02 (shown in the full transcriptome data file of Appendix F.1). This indicates that the addition of pADAP variants to the non-plasmid-bearing wildtype negatively impacts the expression of these chitin-binding proteins(Chapter 7 section 7.2.7). Significant differentiation of expression was also observed between 477 + pA1M02 and wildtype A1M02 (log2fold <2 for *gbpA_1* and log2fold >2 for *gbpA_2*). While the addition of *S. entomophila* pA1MO2 to 477 does not affect the expression of chitin-binding protein 1, the addition of *S. proteamaculans* specific pAfpX and pA1MO2 to 477 shows significant differentiation in chitin-binding protein 2. This shows a negative impact on the transcriptional regulation of chromosomally encoded genes.

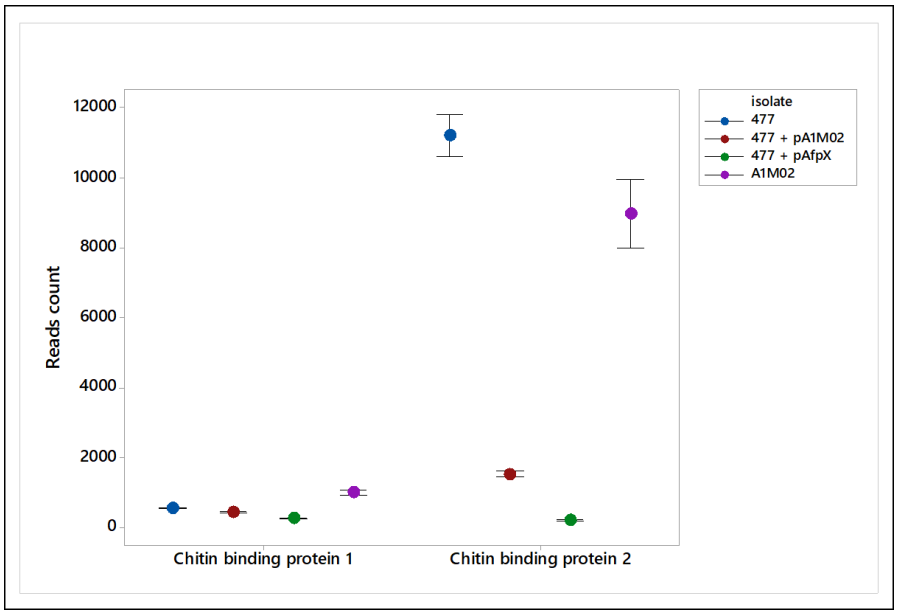


Figure 90 Transcriptome reads count of chitin binding proteins in transconjugants 477 + pAfpX and 477 + pA1M02 compared to wildtypes 477 and A1M02.
Read counts in triplicate. Error bars show standard error of read analysis.

Phospholipases genes identified on isolate 477 were named *phlA* (extracellular phospholipase 1) *pldA*, *pldB* (phospholipase A1 and B) and *tesA* (lysophospholipase L1). Transcriptome analysis of the transconjugants was therefore focused on these four phospholipases. In both transconjugants (+pA1M02 and +pAfpX), the production of phospholipases was downregulated relative to the wildtype

isolate 477 (Figure 91) for the production of extracellular phospholipase A1. In 477 + pA1M02, neither *pldA* nor *pldB* showed any significant differentiation of expression to the wildtype. For *tesA*, the transconjugant 477 + pA1M02 had significant upregulation (log2fold <2) and in 477 +pAfpX; *tesA, pldA* and *pldB* showed a significant upregulation to log2fold <2 compared to wildtype 477.

Comparisons of wildtype A1M02 with 477 + pA1M02 transconjugant show that a significance of differentiated gene expression (log2fold <2) was observed in upregulation of *pldA* in the wildtype, whereas no significance was observed between *tesA, phlA* and *pldB*. This implies that in the instance of 477 + pA1M02, the change in expression relative to wildtype 477 is closer to the plasmid donor chromosome. Downregulation of *pldA* in 477 + pA1M02 in comparison to wildtype A1M02 implies the chromosomal production of extracellular phospholipase A1 is impacted by the addition of a non-native plasmid type.

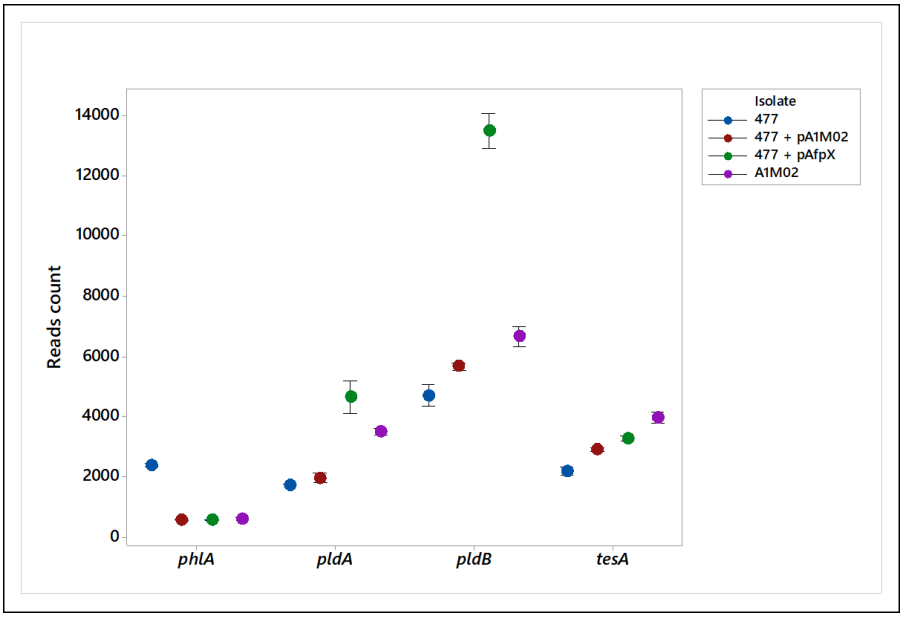


Figure 91 Transcriptome reads count of phospholipase genes in transconjugants 477 + pAfpX and 477 + pA1M02 compared to wildtypes 477 and A1M02.
Read counts in triplicate. Error bars show standard error of read analysis.

Metalloproteases of isolate 477 were searched for in outputs of transcriptome data relative to the transconjugants (+pA1M02 and +pAfpX) (Figure 92). *ftsH,* encoding an ATP-dependent zinc-metalloprotease, was shown to be upregulated in 477 + pAfpX (log2fold change -0.4), whereas the insertion of pA1M02 showed no statistical difference of *ftsH* expression relative to wildtype 477. A similar trend is observed with metalloprotease *pmbA*, where no differences were observed in 477 + pA1M02 when compared with 477, but 477 + pAfpX shows a 2logfold significance of -0.39. This indicates that the presence of an *S. entomophila* type pADAP plasmid in 477 does not affect the

expression of a metalloprotease, but the addition of pAfpX plasmid variant significantly increases its expression (Figure 92).

In wildtype plasmid bearing A1M02 (pA1MO2) compared to 477 +pA1M02, *ftsH* is upregulated (log2fold >2), whereas no statistical differentiation is observed with *pmbA.* This shows that increased expression of *ftsH* is not conferred with the acquisition of the plasmid type, as no difference was shown to wildtype isolate 477. No significant difference in the differentiation of *pmbA* was noted, which shows that the expression of this metalloprotease is not altered by the addition of a non-native plasmid type. As discussed in Chapter 4, isolate 477 and A1M02 both encode for three extracellular proteases defined as protease one, two, and three. Assessment of the expression of these genes shows that, against wildtype 477, 477 + pAfpX is downregulated for both extracellular protease two (log2fold >2) and extracellular protease three (log2fold <2). For 477 + pA1M02, only extracellular protease two showed any significant differentiation (log2fold <2). No change was shown for any transconjugant with extracellular protease three.

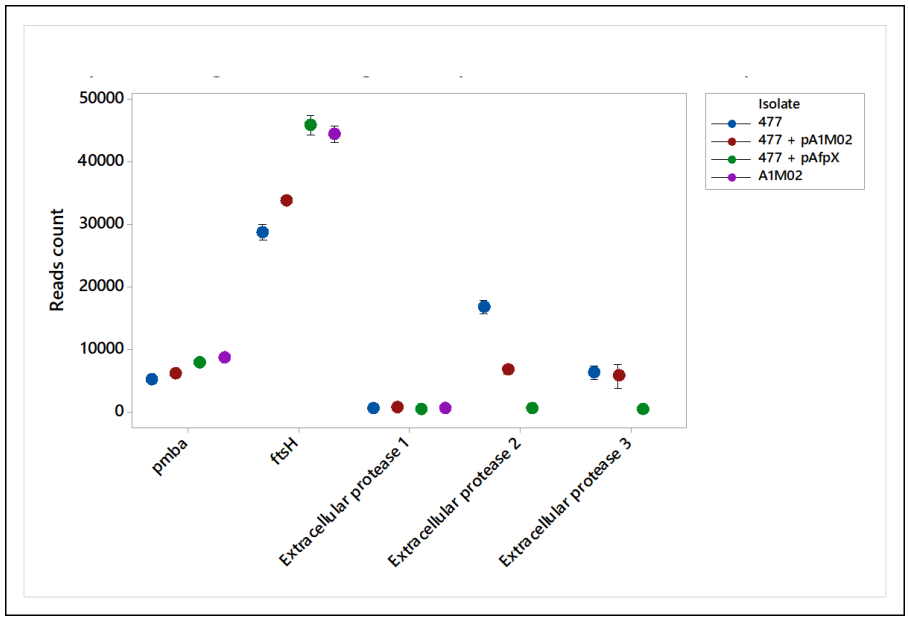


Figure 92 Transcriptome reads count of metalloprotease and extracellular proteases genes in transconjugants 477 + pAfpX and 477 + pA1M02 compared to wildtypes 477 and A1M02.
Read counts in triplicate. Error bars show standard error of read analysis. Extracellular proteases without values for A1M02 could not be shown as confirmed gene homologs.

The addition of either plasmid (pAfpX or pA1M02) results in a negative impact on the expression of accessory virulence determinants in wildtype plasmid-free isolate 477. This corroborates the findings to be reported in Chapter 7, where the addition of pADAP variants conferred inconsistent disease phenotypes.

8.2.2 Plasmid gene regulation

Comparisons were undertaken between the transcripts of pADAP plasmid wildtype A1M02 against transconjugant 477 + pA1M02 to determine if there was any observable effect in the expression of plasmid genes. This comparison analysis was chosen due to the similarities of the chromosomes and the identical plasmid in each. The experimental design meant the transcriptome output should easily identify the pADAP plasmid genes and any changes in the expression of toxin-associated factors such as *sepABC* complex and the Afp. As was undertaken previously, PCA analysis was used to show the clustering of the triplicate sequence reads (Figure 93). Panel B shows that in comparison to the native plasmid bearing A1M02, 477 + pA1M02 shows a high degree of gene differentiation of expression. This analysis was then taken and cross-referenced to the genes of pADAP.

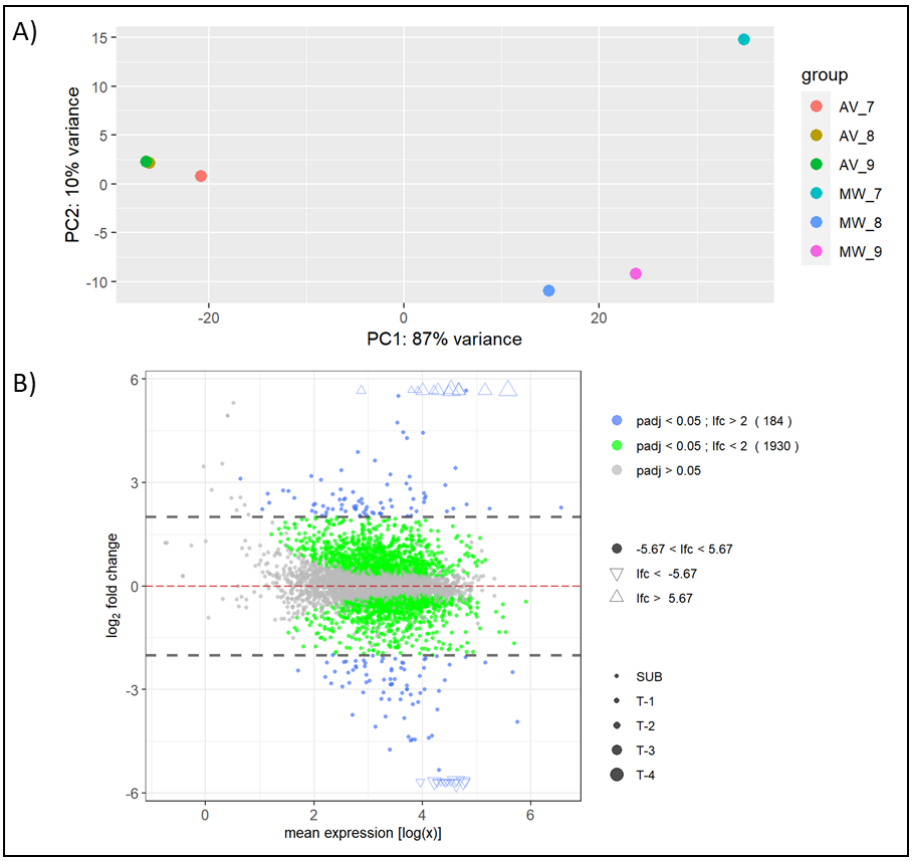


Figure 93 Analysis of transcriptome data in triplicate comprising samples MW7-9 (WT A1M02) and AV7-9 (477 + pA1M02)
A) Principal component analysis of A1M02 and 477 + pA1M02 library of normalised data. B) Transcriptome data analysis between A1M02 and 477 +pA1M02, showing differential gene expression between the transconjugant A1M02 + pA1M02 and the wildtype 477. Blue dots denote genes with significant changes in expression with an adjusted P value cut off 0.05, log2fold change more than 2. Green dots denote significant changes (P>0.05) of genes with log2fold less than 2. Grey dots show genes with no significance. Downregulated genes in the transconjugant are shown below the red line and upregulated genes in the transconjugant above the red line.

To determine any effects on the plasmid gene expression, transcript IDs were compared with the plasmid annotations from assembled isolate A1M02 and cross-referenced to GenBank (accession: NC_002523.5). Full transcriptome data for pADAP plasmid is provided in Appendix F.2.

Table 43 shows the identified transcript number and associated log2fold change observed in the gene cluster for the *sepABC* toxin complex. The addition of the pA1M02 plasmid in 477 increased the expression of the toxin complex, with significant upregulation observed for *sepA* and *sepB* (log2fold <2). No significant change is associated with the expression of *sepC*. The lysozyme 5' of the *sepA* gene also shows a log2fold (<2) increase in relation to the wildtype A1M02 whilst the phage lysis associated endolysin shows increased expression differentiation of log2fold >2. The holin gene shows no difference between the two isolates.

Table 43 SepABC genes differential expression between wildtype A1M02 plasmid donor and transconjugant 477 + pA1M02.
Annotations from pADAP from GenBank cross-referenced with Sitter (2020) annotations.

Protein	Annotation	477+ pA1M02		A1M02		LOG2FC
JJLEJPGA_03778	Holin	139	±13.89	131	±61.26	0.29
JJLEJPGA_03779	Lysozyme	137	11.26±	36	±11.53	**-2.17**
JJLEJPGA_03780	Lysozyme	213.33	± 42.09	385.66	± 80.82	-1.15
JJLEJPGA_03781	SepA	11463.67	± 1133.52	8613.66	± 2300.83	-0.72
JJLEJPGA_03782	SepB	7549	± 775.10	5554.66	± 1354.21	-0.74
JJLEJPGA_03783	DUF2514 family protein	1547.66	± 196.38	1467	± 277.77	-0.38
JJLEJPGA_03784	SepC	7684.33	± 468.01	7879	± 1746.36	-0.27

Bolded shows P >0.01 log2fold >2
Underlined shows P >0.05 log2fold >2

In addition to the *sepABC* toxin complex, investigations were undertaken on the pADAP encoded *afp1-18* genes. As can be seen in Table 44, most of the *afp 1-18* are upregulated in the 477 + pA1M02 transconjugant by a log2fold increase >2. Afp13 and Afp1 showed a log2fold increase (<2) in the transconjugant. These results show that the expression of Afp is upregulated in transconjugant 477 + pA1M02. The mismatch of up and down regulation of individual genes could explain the results of the aberrant bioassay of transconjugants in Chapter 7 (section 7.2.3). In addition to changes in expression of Afp genes, the lysis gene associated lysozyme (*rrrD*) 5' of *afp1* shows a log2fold (>2) differentiation of higher gene expression relative to the wildtype.

Table 44 *Afp* genes differential expression between transconjugant 477 + pA1M02 and wildtype A1M02 plasmid donor.

Annotations from pADAP from GenBank cross-referenced with Sitter (2020) annotations.

Transcript ID	Annotation	477 + pA1M02		A1M02		LOG2FC
JJLEJPGA_03791	Lysozyme *mur1*	23	±6.56	5.67	±4.93	**-2.40**
JJLEJPGA_03793	Afp1	119.67	±16.77	81	±15.62	-0.86
JJLEJPGA_03793	Afp2	2538.66	± 223.28	746	± 165.50	**-2.07**
JJLEJPGA_03794	Afp3	82462.33	± 14421.19	1213	± 302.22	**-6.39**
JJLEJPGA_03795	Afp4	54530.33	± 12685.73	579	± 209.06	**-6.86**
JJLEJPGA_03796	Afp5	6396.66	± 1471.4	171	± 16.09	**-5.50**
JJLEJPGA_03797	Afp6	1346	± 397.05	5.66	± 4.16	**-8.18**
JJLEJPGA_03798	Afp7	11142	± 3849.99	79	± 32.92	**-7.40**
JJLEJPGA_03799	Afp8	29071	± 11092.33	276	± 54.28	**-7.02**
JJLEJPGA_03800	Afp9	14806	± 3998.46	112.66	± 4.93	**-7.30**
JJLEJPGA_03801	Afp10	17787.67	± 4123.50	418.66	± 120.44	**-5.73**
JJLEJPGA_03802	Afp11	60666.33	± 16701.89	373.66	± 78.14	**-7.61**
JJLEJPGA_03803	Afp12	83096.67	± 19656.25	1716	± 338.51	**-5.89**
JJLEJPGA_03804	Afp13	681.6667	± 50.50	218.66	± 24.94	-1.90
JJLEJPGA_04489	Afp14	85543.67	± 5141.35	90931.33	± 19820.70	**-4.73**
JJLEJPGA_04490	Afp15	8875.667	± 489.60	560.3333	± 81.98	**-4.28**
JJLEJPGA_04491	Afp16	2239	± 244.69	222.6667	± 31.66	**-3.62**
JJLEJPGA_04492	Afp17	1096	± 212.58	93.33333	± 32.08	**-3.87**
JJLEJPGA_04493	Afp18	7421	± 382.51	5505.333	± 495.19	**-0.72**

Bolded shows P >0.01 log2fold >2

Underlined shows P >0.05 log2fold >2

Final searches of the plasmid-related genes showed downregulation of additional accessory determinants located on the plasmid associated with pathogenicity. Genes associated with fimbriae and pilus production were noted to be significantly down-regulated in their expression in the transconjugants (full table can be found in Appendix Table F 1). These genes have been hypothesised to be implicated in pathogenicity as potential host colonisation appendages.

In summary, the inconsistency in detected disease when larvae were treated with transconjugants reported in Chapter 7 is consistent with the results found in this analysis. The combined effect of upregulation of some virulence factors whilst downregulation of others observed could result in an erratic disease pattern that struggles to establish a coherent infection. The reduction in overall gene expression of accessory virulence regions implicated in host colonisation again could be implicated in the low numbers of bacteria re-isolated from grass grub larvae in bio-infectivity assays (Chapter 7 section 7.2.5). This could be in part due to the reduced capacity for virulence and therefore establishment relative to the wildtype.

8.3 Discussion

Assessing the transcriptome of selected transconjugants and the wildtype plasmid donor allowed for further investigation into the results seen in Chapter 7. These results support the hypothesis of this study and by Sitter (2020) that the chromosome and plasmid have coevolved with a precise regulatory relationship. The downregulation of expression of accessory virulence determinants could have implications relating to the erratic nature of the less virulent disease state seen in transconjugants (Chapter 7).

Comparisons between accessory virulence determinant expression relative to the plate assays conducted in Chapters 4 and 7 indicate that, in some instances, the conjugation of pADAP into other chromosomal backgrounds confers the level of production seen in the wildtype pathogenic isolate. In the case of chitinase production, 477 + pAfpX is indicative of an increased burden on the host cell. The DNase encoding gene *nucA* shows an interesting trend in expression that can be related to plate assays of wildtype isolates in Chapter 4 as well as to transconjugants of Chapter 7. As previously discussed, DNase plate characterisation showed a significant difference between pathotypes, which seems to be conferred with the addition of pADAP. Association of downregulation to a transcriptional regulator, where regulation of *nucA* of *Serratia marcescens* assigned to *nucC* was described by Jin et al. (1996), showed that no consistent trend could be found in the transconjugants. In this instance, only 477 + pA1M02 was downregulated relative to the wildtype.

Chitinase expression was also found to be negatively impacted by the addition of pADAP in both +pA1M02 and +pAfpX transconjugants, as well as one of two chromosomally located chitin-binding proteins. Chitin binding proteins of *Bacillus thuringiensis* have been implicated in early infection as well as in adhesion in the peritrophic matrix of *Galleria mellonella* (Qin et al. 2020). If this is also true in the gut of *C. giveni* larvae (no bacterial binding site has yet been identified) the downregulation of one of these genes could lead to the overall impediment of the initial infection phase and may contribute to the aberrant bioassay data observed with the transconjugants (section 7.2.3-7.2.4).

Changes were also observed in the expression of metalloprotease and extracellular proteases. The addition of pADAP (pA1M02) to isolate 477 confers a similar level of pathogenicity to wildtype plasmid-bearing isolates, so it could be assumed that expression of these genes would either not change or increase. What was observed however was that (most notably in 477 + pAfpX) the addition of the non-native pADAP plasmid to the wildtype plasmid-free isolate 477 resulted in the downregulation in two of the three extracellular proteases encoded on the chromosome. Interestingly, with the 477 + pA1M02 transconjugant, differentiation of expression levels was only seen in extracellular protease two. It can be hypothesised that the coevolution of the pADAP plasmid would therefore allow it to

confer a less debilitating effect on the host cell, reducing the degree of burden on gene expression. This scenario would explain why we have such diverse differences in transcriptome and plate assay effects through the introduction of an alternate plasmid.

In phospholipase A1, a phospholipase implicated in virulence in bacteria, significant differentiation was observed in gene expression. *phl*A1 deficient mutants of *Yersinia enterocolitica* have been shown to induce less inflammation of tissue in murine models, implicating induction of this gene in pathogenesis (Schmiel et al. 1998). Intriguingly, the addition of either of the pADAP variants to wildtype plasmid-free isolate 477 negatively impacts the expression of this gene. This is again found on plate assays in Chapter 7, section 7.2.7. The addition of the plasmid confers a disadvantage to the host cell that impedes the production of some accessory virulence determinants. As previously determined in Chapter 4, no statistical difference was found between pathotypes of wildtype isolates of either *S. proteamaculans* or *S. entomophila* relative to the production of selected accessory virulence determinants on selective media except for DNases.

Buckner et al. (2018) found that the acquisition of plasmid encoding antimicrobial resistance in *Klebsiella pneumoniae* resulted in significant changes to chromosomal and plasmid expression of genes. It was theorised this was to allow the bacterial plasmid to acclimatise to the chromosome and reduce the associated cost of carrying the plasmid. The resulting plasmid-bearing isolates were shown to have a reduced ability to form a biofilm, another accessory virulence factor. This corroborates what was found in this study where the acquisition of pADAP reduced the expression of accessory virulence-associated genes in *S. entomophila* isolate 477. In addition to chromosomally encoded genes and regions that are impacted by the presence of a non-native plasmid, the expression of pADAP associated genes in 477 + pA1M02 were altered. Upregulation of the lysis genes 5' of both the Afp and *sep* regions, which are proposed to enable the release of Afp/Sep, implies that potentially uncontrollable cell lysis is occurring, which would be detrimental to the cell. The upregulation seen in these genes combined with the downregulation of accessory virulence factors specifically the Sep-TC could result in the reduction of a bacterial colony's ability to establish and proliferate. This reduction in cell proliferation was shown in Chapter 7 (section 7.2.5) where transconjugants showed a reduced re-isolation CFU level in comparison to the wildtype bacteria supports this hypothesis. Greater lysis might explain less isolation of cells and the observed smaller colony sizes observed in the 477 + pA1M02 transconjugants and would negate any benefits from the overexpression of the insecticidal *sep*-TC and Afp genes (section 8.2.2).

Modi and Adams (1991), in assessing *E. coli* and its derivative plasmids, suggested that in the context of plasmid bacterial evolution five evolutionary responses could be recorded. In one of these

responses, adaptive genetic changes noted in the chromosome was shown to be independent of plasmid presence or control. In the context of this study, there is potential to apply this evolutionary model to non-pathogenic isolate 477 with the addition of pADAP from A1M02. As discussed above, the genes associated with notable accessory virulence determinants were downregulated, whilst lysis genes and secondary virulence determinants on the plasmid were upregulated. The downregulation of these chromosomally encoded genes could infer that the chromosome has adapted to the increased burden presented by the presence of the plasmid. This is, again, supported by transconjugant bioassays where the pathogenicity is conferred by the plasmid, but to a lesser degree. This implies that, specifically with pADAP from *S. entomophila*, conjugation can confer comparative disease, but as found here, reduces the capacity of the cell for expression of important genes encoding enzymes or metabolic gene clusters.

Overall, the hypothesis that the plasmid and cell have coevolved is supported by the outcome of this transcriptomic analysis. This is seen for the plasmid-encoded genes as well where the presence of a non-native plasmid alters the expression and increases the lysis of the TC and Afp, further supporting *in vivo* and *in vitro* findings of Chapter 7. Further investigation to expand on experimental design would be warranted, where limitations in growth medium and growth phase are recognised in this study. *In vivo* transcriptome analysis would allow for the investigation of secretion systems activated by the presence of eukaryotic cells (type 3 and type 6 secretion systems). In addition, analysis at different stages of bacterial growth would allow for a more complex and overall picture of the secretome. Growth phase differences in the global transcriptome of *Helicobacter pylori* showed significant change in the expression of virulence-associated genes at the late log to stationary phase transition (Thompson et al. 2003). Further analysis of the *Serratia* transconjugants therefore may reveal relevant changes in the virulence-associated transcriptome.

Chapter 9

General Discussion

In this thesis, the chromosomes of both pathogenic (amber disease-causing) and non-pathogenic (generally plasmid-free conspecifics) of two species of *Serratia*, were compared to determine chromosomally bound features that could facilitate disease or plasmid persistence within the population. Initially, several hypotheses were proposed (Chapter 1, section 1.10). It was originally hypothesised that the *S. entomophila* bacterium is genetically conserved, where metabolic and regulatory clusters promote the chronic disease state. In addition, both *S. entomophila* and *S. proteamaculans* have coevolved with pADAP, and finally, that hypervirulent isolates show increased production of accessory virulence determinants compared to chronic pathogens. Overall, the results showed that the content of the chromosomes of *S. entomophila* are generally conserved, with regions of differentiation aligning to putative genomic islands and DNA acquisition events. It is suggested that both species have coevolved with pADAP in New Zealand, but unidentified potential regulatory mechanisms mediate hypervirulence in *S. proteamaculans* AGR96X.

9.1 Is *S. entomophila* genetically conversed compared to *S. proteamaculans*?

To address the hypotheses stated in Chapter 1, that *S. entomophila* is a genetically conserved bacterium as opposed *to S. proteamaculans*, as implicated by the variable pathotype, full genomic comparisons between the two species were conducted. Through Chapters 4, 5, and 6, *in silico* and *in vitro* results demonstrated the conserved chromosome of *S. entomophila* and the extent of variation observed. These full comparisons are summarised below.

9.1.1 Is *S. entomophila* undergoing genome reduction?

One of the main hypotheses of this thesis was that *S. entomophila* chromosomes are undergoing genome reduction in response to long-term bacterial-grass grub associations. If correct, *in silico* observations should show the loss of non-essential gene regions that would not benefit the bacteria's survival in the host-pathogen stasis. By undergoing whole-genome comparisons of *S. entomophila* and *S. proteamaculans*, the evidence was then generated to decipher as to whether this was occurring. In marine free-living bacteria it was hypothesised that increasing their metabolic efficiency by utilising substrates readily in the environment could advance their evolution by reducing the genome as a result of selection for the loss of redundant synthesis pathways in a process of trophic specialisation (Qin et al. 2019). When comparing *S. entomophila* to *Serratia symbiotica* for the genomic features of a

shrinking chromosome, there is limited evidence that this is occurring (Manzano-Marín and Latorre 2016). Though a reduced number of metabolic pathways, reduction in predicted genomic islands was recorded the bounds of the pangenome were generally calculated as open. The reduced number of unique genes present per *S. entomophila* isolate suggests that, relative to *S. proteamaculans,* the *S. entomophila* chromosome has reduced plasticity.

The overall chromosome size, although smaller than *S. proteamaculans,* is proportionate to the number of coding regions. Across both species, an average coding density of 87-88% was recorded showing an even coding density and suggesting no differences in pseudogenes between *S. entomophila* and *S. proteamaculans. Serratia symbiotica,* a symbiont of aphids, has previously been studied as a model organism for genome reduction, meeting the criteria listed previously wherein genome size and gene loss in biosynthetic pathways resulted in a smaller genome. As a symbiont, *S. symbiotica* does not need the ability to convert multiple types of nutrients, therefore, can afford the loss of some of these biosynthetic pathways. A study conducted by Manzano-Marín and Latorre (2016) found that by comparing the coding regions of specific proteins in *S. symbiotica* isolates to those of *E. coli* K12 and *S. marcescens* DB11, non-essential biotin/lypoyl protein domain loss (deemed non-essential by studies in *E. coli*) were observed. To again confirm that *S. entomophila* is not undergoing genome reduction, this aspect of the Manzano-Marín and Latorre (2016) study was emulated where the assessment of *S. entomophila* isolates revealed that these non-essential domains were as prevalent across the whole cohort. Another characteristic of a genome undergoing reduction is the number of non-coding RNAs such as tRNAs and rRNAs. Previous studies comparing the prediction of ncRNA families in *S. proteamaculans* and *S. symbiotica* show that early stages of genome reduction display as reduced numbers of ncRNA, with the downstream protein-coding gene remaining intact (Matelska et al. 2016). Comparing the encoded copies of the individual genes and their operons between the bacterial *Serratia* species included in this study shows there is no overall ubiquitous reduction in the presence of tRNA and rRNA sequences.

In conclusion, there is no evidence in support of genome reduction having occurred in *S. entomophila.*

9.1.2 Species-specific gene acquisition

The presence of the itaconate degradation cluster in *S. entomophila* is a specific and useful example of species-specific acquisition, with potential impacts on bacterial fitness. It is important to note that no other member of the *Serratia* genus has been found to carry the itaconate degradation operon. Comparisons between the *Serratia* spp. and predictions of genomic islands, and its inverse orientation in several isolates suggest that this itaconate mobile element was acquired only by *S. entomophila*, with closest similarities to the *Yersinia pestis* three-gene operon described by Sasikaran et al. (2014).

Neither orientation of this operon correlated with a particular sub-group of *S. entomophila*. Investigations in this study found that this operon likely confers a metabolic advantage to *S. entomophila*, elucidated by knockout experiments of the *mcl1* gene. We can, however, postulate that this itaconate degradation cluster replaces this function in other *Serratia* that instead encode for an additional carbohydrate metabolism pathway (the De Ley Doudoroff pathway) that results in the same output of pyruvate. As found by Hossain et al. (2019) the deletion of the homologous gene *tmtA* (in this study described as *yfdE*) resulted in the cessation of itaconic acid production. Further study, therefore in knocking out the *yfdE* gene in *S. entomophila* would be beneficial for future research to confirm its role, although Sasikaran et al. (2014) found that not all itaconate degradation gene mutants produced the same effect of survival in macrophages or metabolic benefit.

Based on a study by Pujol et al. (2005), preliminary investigations were undertaken to ascertain any potential function of this itaconate pathway for survival in host amoeba, analogous to the requirements of *ripA* (*mcl1* ortholog) in *Yersinia pestis* for its replication in macrophages. Experimental results with bacterial mutants indicated that this was not the case in lysates of amoeba samples from larval macerates. Further investigation using 'clean' cultured amoeba samples may be needed to further explore this hypothesis. Detection of an itaconic acid-producing organism that colonises the soil or host, showing a benefit to *S. entomophila* bacterial isolates in terms of itaconate utilisation, would solidify the findings of this study that *S. entomophila* is undergoing trophic specialisation to its environment.

In vivo grass grub competition assays of Mcl1::tet mutant against wildtype 626 showed that over twelve days, the endpoint re-isolation CFU's showed no statistical difference, though a significant difference was observed at day three indicates that the mutant required a longer time to establish in the host. However, additional work is required to further characterise the itaconate degradation operon, with comprehensive operon knockouts of *mch* (*ripB* orthologue) and *yfdE* (as previously discussed), alongside complementary *in vivo* and *in vitro* testing to determine the precise role of this pathway in the lifestyle of the bacterium. Assessing changes in carbon assimilation of the wildtype and mutant by OmniLog would further fortify this data by quantifying the full cost/ benefit ratio.

In short, analysis of the itaconate degradation cluster in *S. entomophila* isolates may support the hypothesis (iii) that metabolic clusters found in *S. entomophila* promote the chronic nature of the bacterium. Further investigation would need to be undertaken to confirm or reject this support in relation to pathogenicity, but preliminary assays show that an observable trend of metabolic advantage is conferred with the addition of the cluster. It is however plausible that this region may reflect lifestyle speciation of *S. entomophila* with its host.

9.2 Do metabolic clusters in *S. entomophila* aid maintenance of a chronic disease state?

Concerning hypothesis ii (Chapter 1, section 1.10), no evidence was found to support that metabolic and gene regulatory clusters in *S. entomophila* could promote the maintenance of a chronic disease state. However, it was shown that the encoded metabolic clusters in *S. entomophila* are more uniform across pathotypes than those in *S. proteamaculans*. For *S. proteamaculans*, the presence of a pyrrolnitrin (an antifungal antibiotic) described in Chapter 6 encoded on some isolates of both subspecies of *S. proteamaculans* (*quinovora* and *proteamaculan*s) is a good example of the variation seen in the encoded metabolite pathways. Furthermore, plasmids in their natural host isolate of *S. entomophila* exhibit a consistent disease phenotype as opposed to the variability of *S. proteamaculans*. The potential exists that the homogeneity in encoded metabolic modules could assist the fitness of the cell, resulting in a consistent chronic infection.

Comparisons of predicted biosynthesis and amino acid biosynthesis modules can also be deemed an indicator of co-existing symbionts. Analysis of intact encoded metabolic pathways revealed that *S. entomophila,* in comparison to *S. proteamaculans,* showed incomplete modular pathways for the synthesis of pyruvate via the De Ley Doudoroff pathway. As discussed previously in Chapter 5, fungal species (*Aspergillus* and *Ustilago*), regularly found in soil samples, are known to produce itaconic acid, therefore it can be suggested that *S. entomophila* has the opportunity to utilise environmental itaconic acid as a source for carbohydrate metabolism, negating a more laborious and energy-consuming means. The complementation of metabolic pathways in bacteria, with syntrophic interactions between microbial communities, has been previously described as the most costly amino acid metabolism pathways (Mee et al. 2014). In *C. giveni*, fungal pathogens also occur, primarily *Metarhizium* spp. and *Beauveria* spp. It would therefore be useful to see if these known fungal pathogens produce itaconic acid, thereby providing *S. entomophila* an energy source. It is not beyond the scope of bacterial synergy that the utilisation of various secondary compounds from other soil-dwelling organisms could be utilised by *S. entomophila* to reduce metabolic burden. Various organisms have been described as bacterial cheats, benefiting from environmental metabolites or products synthesised by other bacteria with no cooperation in return (Bruce et al. 2017). This theory of bacterial cheats is currently being explored in the *Serratia* spp. in research aligned to this study (Connor Watson, personal communication). Further analysis of metabolic systems would be beneficial, with the exploration of the transcriptome data comparing the two species and variable pathotypes potentially underlining the mechanisms for *S. entomophila* to hold a more chronic disease state than *S. proteamaculans*.

9.3 Do chromosomally encoded genes have a role in disease?

In line with the aim to distinguish differential expression in hypervirulent isolates, the findings of accessory virulence determinants included the production of various enzymatic compounds by the bacteria (e.g. DNases, proteases, chitinases). This was used to determine whether the differences identified *in silico* corroborate the findings *in vitro*. Interestingly, the main species differentiation character between *S. entomophila* and *S. proteamaculans* (Chapter 3) concerned the production of extracellular nucleases. DNases produced by *S. marcescens* are already established as defining characteristics useful in distinguishing *S. marcescens* from other Yersiniaceae (Black et al. 1971, Benedik and Strych 1998). The extracellular nuclease *nucA* defining *S. marcescens* as a DNase-producing organism is uniformly found across isolates of both *S. entomophila* and *S. proteamaculans*. The *nucC* gene has been implicated as a transcriptional regulator of *nucA* in *S. marcescens* and via the SOS response (Jin et al. 1996). However, research by Grkovic and Mahanty (1996) found that a *recA* mutant lacked the ability to produce the *S. entomophila* extracellular nuclease (DNase). The variable production of extracellular nuclease *nucA*, and its association with the SOS regulator *recA,* could explain the inconsistencies seen between *S. entomophila* and *S. proteamaculans*, whereby the latter has significantly reduced nuclease production. The fact that both species encode the same extracellular nucleases suggests that the regulation of these differs, having potential implications on horizontal gene transfer in high DNase-producing *S. entomophila.*

As was discussed in Chapters 4, 5, and 6, the expression of accessory virulence factors such as chitinases and phospholipases can be involved in invasion techniques as host tissue degrading enzymes or adhesins. Though no significant differences could be seen on plate assays for the production of these accessory virulence determinants between chronic and hypervirulent pathogens, a slight trend was observed that was accentuated in plasmid transconjugants (Chapter 7). In the assessment of plate assays, siderophore production by hypervirulent isolates was significantly elevated more than in either chronic or non-pathogenic isolates. Further *in silico* analysis showed that each of the *Serratia* isolates assessed encodes a siderophore, with no unique siderophore encoding region identified that would implicate hypervirulent isolates in a unique mechanism. What this confirms is that an as yet unidentified underlying regulatory mechanism modulates the use of siderophores as iron sequestering molecules, benefiting hypervirulent isolates. Specifically, hypervirulent isolate AGR96X shows the highest average identification of siderophore production using plate assays of all *S. entomophila* and *S. proteamaculans* assessed. Overproduction of siderophores has been associated previously with hypervirulence in gram-negative bacteria, with the characterisation of hypervirulent isolates of *Klebsiella pneumoniae* that secrete more iron sequestering molecules than isolates related to the classical infection (Russo et al. 2011). Noted increase of pathogenic potential in other species solidifies the findings of this study, that the potential for virulence increases with the over-production and

potential increased activity of iron-chelating molecules. Further research could be conducted to quantify these results and establish whether the siderophores of hypervirulent isolates do indeed have increased biological activity. The possibility, therefore, is that in addition to the presence or absence of additional genes associated with virulence or nutrient sequestering, that presence alone of the genes and their regulatory factors does not solely confer a benefit.

In conclusion, the presence of chromosomally encoded genes in *S. entomophila* supports hypothesis i (Chapter 1, section 1.10) that states that *S. entomophila* is genetically conserved. However, there is limited evidence to support hypothesis v (Chapter 1, section 1.10) that hypervirulent strains encode more chromosomal bound factors that promote virulence or fitness. Further quantification of regulatory factors and investigation into siderophore production of hypervirulent *S. proteamaculans* isolates could further elucidate whether underlying mechanisms exist to contribute to virulence and fitness.

9.4 Where is disease evolution heading, and what is the role of HGT?

From the comparative genomic analysis, there is little to indicate that *S. entomophila* is undergoing genome reduction. *S. entomophila* has regions of horizontally acquired genomic material such as the itaconate degradation operon that are not found in other *Serratia* spp. However, *S. entomophila* does show a lower rate of genomic inversion or rearrangements when compared to *S. proteamaculans* which may relate to a reduced prediction of IS elements found in *S. entomophila*. This shows again that the chromosome of *S. entomophila* is more genetically conserved than that of *S. proteamaculans.* Supporting the *in silico* analysis, bacterial BOX fingerprinting in Chapter 4 also shows that the profile of *S. entomophila* is less varied than *S. proteamaculans*. There is the possibility that the reduced evidence for HGT events, will, in the future, lead to a less open genome. Specifically, where fewer IS elements are found, a reduced probability of genomic inversions and deletions also occurs. Currently, the evidence is not yet clear though initial investigations suggest differences. The variation observed in the phenotype and pathotype of *S. proteamaculans* isolates is in part due to a diverse range of encoded biosynthetic compounds that are likely beneficial to colonising an ecological niche. As these clusters are more varied in *S. proteamaculans* than in *S. entomophila*, the 'clonality' of the putative pathogens found in the *S. entomophila* species can be attributed to the relative similarity in its encoded accessory genes and biosynthetic compounds. The varied encoded accessory genes in the more genetically diverse S*. proteamaculans* is a phenomenon more likely due to the frequent and diverse nature of the chromosome in this species.

In silico analysis of the chromosomes of both *S. entomophila* and *S. proteamaculans* revealed that *S. proteamaculans* have higher levels of potential HGT events. Overall, there are many differences observed between the two amber-disease causal agents. As discussed in Chapter 5, specifically in relation to *S. entomophila,* transposases appear less abundant in various insert positions on the chromosome, implying fewer horizontal gene transfer of transposases and more vertical transmission of DNA. Subsequently, more individual transposases were identified in *S. proteamaculans* isolates than *S. entomophila*. The reduced number of encoded IS elements also reduces the opportunity for mutations that facilitate genomic rearrangements such as inversions or deletions. This, together with the reduced number of predicted genomic islands alongside the *in silico* detection of extracellular DNases and low-level openness to the pangenome of *S. entomophila,* suggests there may only be limited acquisition of new DNA by this species. Genomic assessments of *S. proteamaculans* showed other defining features such as higher levels of encoded genomic islands and transposases in randomised genome insertion points, making this organism more acclimated to the acquisition of new genetic material.

9.5 Do chromosomes select for plasmids?

To address the overarching aim of this study to determine if the chromosomes *S. entomophila* have coevolved with *C. giveni* and to assess variability of *S. proteamaculans*, transconjugants were constructed and assessed *in vitro and in vivo* (Chapter 7) and examined through preliminary transcriptome data (Chapter 8). The results supported hypothesis iv (Chapter 1, section 1.10). Overall, the addition of variable pADAP plasmid types to the chromosomes of alternate isolates did confer pathogenicity, as was expected due to previous studies by Sitter (2020) and Grkovic et al. (1995). It was observed that the addition of variable pADAP plasmid types has a negative effect on the production of accessory virulence determinants. Analysis of the transcriptome outputs of the so-termed accessory virulence genes identified in Chapters 5 and also supports the hypothesis that the plasmid and chromosome of *Serratia* spp. have coevolved. This is complimented by the variable results found *in vivo,* where pathotypes showing mixed inducibility after the conjugation of the pAfpX, a phenomenon that was not related to loss of viability of the isolate in the larvae. These results imply that pADAP has likely coevolved with its chromosome. In certain plasmid families such as the *Bacillus thuringiensis* (BT) plasmid families, promiscuous plasmids are defined as plasmids that can colonise a broad range of host cells, self-conjugating and stably replicating in at least two subgroups of Proteobacteria (Li et al. 2015). It was observed in *Serratia* that pADAP can be conjugated into *S. entomophila* and *S. proteamaculans,* demonstrated in a laboratory environment. However, data presented in this thesis reveals the full potential of the plasmid-encoded disease is not reached for all plasmid-chromosomal combinations.

Results of the assessment of transconjugant support the conclusion that initiation of the disease phase is mediated by an underlying regulatory mechanism that is not conferred with the addition of pADAP. This was also inferred in the experimental investigation of the SOS response induction of virulence factors in isolates of *S. entomophila* and *S. proteamaculans*. It was observed that, of the assessed isolates, AGR96X was the only consistently inducible isolate with MitC over the three years of bioassay assessments. It was determined that AGR96X encoded either a LexA-RecA dependent or independent pathway to promote the overproduction of AfpX that is not present in other assessed isolates. Various candidates were identified including LysR family regulators, though further analysis utilising transcriptome analysis of induced vs uninduced cultures currently underway (Mark Hurst, unpublished data) seeks to further understand this mechanism.

Transcriptome analysis was able to be undertaken on transconjugants 477 + pAfpX and +pA1M02 with wildtype pathogenic A1M02 (+pA1MO2) and wildtype plasmid-free 477. Comparisons were made between wildtype A1M02 and 477 + pA1M02 for expression differentiation related to the pA1M02 (pADAP) plasmid. Differentiation in gene expression found *in silico* (Chapters 5 and 6) and linking to plate assays (Chapter 7), show that the addition of pADAP plasmids downregulated expression of accessory virulence genes in the transconjugant relative to the wildtype 477. This again supports the hypothesis that the plasmid and chromosome have coevolved, to allow the cell to negate the burden of carrying the plasmid. When assessing pA1M02 associated *sepABC* and *afp* genes, these regions and more so the upstream lysis genes are all upregulated in 477 + pA1M02. Though *sepA* and *sepB* are upregulated in comparison to wildtype A1M02, the difference between the two is low. In the case of *sepC* (the effector gene), no difference is seen between the wildtype and the transconjugant. This indicates that the upstream lysis cassettes are more significantly differentiated between the wildtype and transconjugant 477 + pA1M02. Increased expression of lysis potentially explains the reduction in re-isolation in bio-infectivity assays of transconjugants (Chapter 7, section 7.2.4) and the smaller colony sizes observed in the pAfpX transconjugants (Chapter 7, section 7.2.5). Further investigation would be required to define the relationship more accurately between the plasmid and chromosome, but it is clear that coevolution has occurred to facilitate a balance in gene transcription. Research undertaken parallel to this study by Connor Watson seeks to elucidate the relationship between bacterial cheats and would further indicate plasmid burden and potential advantages to maintaining a plasmid-free population in *S. entomophila.*

9.6 Is there a geographic link to strain occurrence?

The geographic distribution of disease-encoding and non-pathogenic isolates was investigated in this study. Aside from the Chatham Island and West Coast of the South Island, no geographic links could be found among *S. entomophila* isolates. Interestingly, both West Coast isolate MC2 and Chatham Island isolates (440, 442, and 294), though being geographically distant, are both from regions devoid of grass grub. It can be hypothesised that the lack of drivers for evolutionary change e.g. a host-symbiont relationship, reduced the need for these isolates to adapt. Or, Occam's razor would suggest a pathway of transmission potentially between these two regions, such as suggested for the migratory patterns of the white-fronted tern (Aikman and Miskelly 2004). We can postulate the reasons why geographic analysis didn't produce any conclusive links. For example, the application of biocontrol agents utilising *S. entomophila* has been ongoing since the 1990s, and undoubtedly repeated application of isolates of 626 and A1M02 for the best part of 30 years can be assumed to have an overall effect on the distribution patterns of this species. The majority of isolates used in this study were sourced from field sites utilised in the application of these biocontrol measures. Jackson (1990) reported that rooks are accountable for up to 22% of grass grub mortality consistent with earlier reports in Hawkes bay by McLennan and MacMillan (1983). Dense rook populations have since been eliminated in New Zealand, but effects from previous generational avian movement and predation of *C. giveni* larvae could well still be seen in the soil microbiota. Distribution ranges for these birds vary widely, but it is not difficult to extrapolate avian predation of larvae of *C. giveni* would aid in the distribution of bacterial isolates.

9.7 Phylogenetic implications of this study

Phylogenetic assessment of isolated samples of *S. entomophila* from various parts of the world showed that GenBank 16S sequences of isolates from France, Mexico, India, and South Korea fit well into the *S. entomophila* clade, whereas another isolate (CT8) from India appears to be a misidentified *S. marcescens*. It was observed that in 16S sequencing, *S. ficaria* isolate 457, places within *S. entomophila* clade. Further analysis however confirmed that these are genetically distinct species. This reinforces the need for additional methods in tandem with 16S phylogenies to determine species boundaries. For French isolate 220, full genomic comparisons show that genetically this isolate is extremely similar to New Zealand isolates with an ANI of 99.62% to *S. entomophila* isolate 626 (Chapter 5, section 5.2.2). With no large-scale differences found, it is proposed that isolate 220, and potentially other *S. entomophila* isolates, originate from New Zealand and may have been isolated from cross-contamination with studies undertaken between research institutes.

9.8 Conclusion

This study aimed to test various hypotheses that would support the differences in disease phenotypes between *S. proteamaculans* and *S. entomophila* determined by comparative genomics and plasmid transconjugants. While it cannot be disproved that *S. entomophila* is moving towards a conserved genome, evolving in synchronisation with *C. giveni* larvae, the results of this study suggest that this has not occurred to a degree to prove a symbiotic nature. Though *S. entomophila* does not appear to be undergoing genome reduction as noted with symbionts, fewer encoded phage elements and alternate metabolic pathways (itaconate degradation), as well as increased production of nucleases, suggest the bacterium is undergoing speciation with its host. With future work and characterisation building upon the results found in this and previous studies, the mechanisms of regulation and control of the chromosome upon the disease state of *C. giveni* as well as the relationship of *S. entomophila* with its host could then be alluded to. Conclusions suggesting the coevolution of pADAP to *Serratia* spp. serve as a model for chromosome-plasmid interactions. This study demonstrates the selectivity of pADAP to the host chromosome, alluding to pADAP being a more host-restricted plasmid than previously thought.

9.9 Future work

9.9.1 Genomic comparisons of *S. entomophila* and *S. proteamaculans*

As shown in Chapters 4 and 5, the genes encoded on the chromosomes of each isolate referring to the accessory virulence determinants were mostly conserved in *S. entomophila*. It can therefore be concluded the differences lie in a regulatory mechanism. A more extensive analysis of transcriptome data could potentially show the regulation of these genes in wildtype isolates and the transcription of accessory genes in chronic isolates in comparison to non-pathogenic and hypervirulent AGR96X. Transcriptome analysis with transconjugant 3041 + pAfpX could then allow comparisons to wildtype AGR96X (as both are similar organisms), which would help to characterise the burden of the pAfpX plasmid on the expression of pAfpX genes in a non-native chromosome. Analysis of transconjugants of isolate iDIA alongside heat-cured and wildtype pathogenic iDIA would further present a comprehensive analysis of plasmid burden, with regards to a non-native chromosome that already encodes a pADAP plasmid. Comparisons of these transconjugants against both wildtype iDIA and heat-cured iDIA would potentially demonstrate the benefit of carrying a plasmid in this isolate furthering the suggestion that the plasmid and chromosome have coevolved. The ability to then compare the results of comparisons on phenotypic, genomic, and transcriptomic interactions would add depth to the data and a more robust conclusion as to the key drivers of pathogenicity.

Including recent and geographically diverse isolates in further analysis could further categorise the diversity of both *S. entomophila* and *S. proteamaculans*. By including more recent samples a model for evolution could then be built to determine whether *S. entomophila* is moving towards a more endopathogenic relationship with its host. As bioinformatic skills were developed in the process of this study, further comprehensive analysis could not be completed in the time scale that could have been beneficial in the comparison of these two species, therefore further *in silico* analysis would benefit the outcomes of this study. COG analysis between pathogenic and non-pathogenic isolates could also be useful to direct us to any potential differences on the chromosome. Further comparative COG analysis could be undertaken on *S. entomophila* and *S. proteamaculans* to determine any uniform changes between the two species.

Quantitative assays of accessory virulence determinants such as proteases and lipases would enhance information derived from plate assays to more effectively determine production amounts of these accessory virulence factors.

9.9.2 Itaconate degradation cluster as a metabolically beneficial system

Indicative results in this study suggest the itaconate degradation pathway influences wildtype metabolism. However, the use of full operon knockouts would solidify the results of interrupting the *mcl1* gene, alongside full operon *trans* complementation to restore the wildtype phenotypes. Further growth curves in a medium with itaconic acid may also show the effects the deletion of the region has in facilitating survival in environments with increased itaconate.

9.9.3 The geographic diversity of *C. giveni* genotypes

It was noted that there were variable results between years and field sites across the breadth of this study in relation to the susceptibility of grass grub larvae to isolates of *S. entomophila* or *S. proteamaculans*. In previous studies, such as that by Sitter (2020), that the full genome of *C. giveni* has yet to be characterised, meaning there is potential for less susceptible subspecies to coexist in population-dense regions. It would therefore be pertinent that future work would characterise the genome of *C. giveni* larvae from geographically diverse regions, and potentially map the larval subspecies classifications to reduce the variability of results acquired and increase specificity to a specific population that could therefore benefit the development or identification of further potential biocontrol candidates.

References

Acinas, S. G., L. A. Marcelino, V. Klepac-Ceraj and M. F. Polz (2004). "Divergence and Redundancy of 16s Rrna Sequences in Genomes with Multiple Rrn Operons." J Bacteriol **186**(9): 2629.

Aikman, H. and C. Miskelly (2004). Birds of the Chatham Islands, Department of Conservation.

Alberts, B. J., A; Lewis J; et al (2002). Biol Cell. New York, garland Science.

Alikhan, N. F., N. K. Petty, N. L. Ben Zakour and S. A. Beatson (2011). "Blast Ring Image Generator (Brig): Simple Prokaryote Genome Comparisons." BMC Genom **12**: 402.

Anderson, M. T., L. A. Mitchell and H. L. T. Mobley (2017). "Cysteine Biosynthesis Controls *Serratia Marcescens* Phospholipase Activity." J Bacteriol **199**(16): e00159-00117.

Andrews, S. (2010). "Fastqc." 2017.

Angerer, A., B. Klupp and V. Braun (1992). "Iron Transport Systems of *Serratia Marcescens*." J Bacteriol **174**(4): 1378.

Anthony Mason, C. and J. E. Bailey (1989). "Effects of Plasmid Presence on Growth and Enzyme Activity of *Escherichia Coli* Dh5α." Appl Microbiol Biotechnol **32**(1): 54-60.

Apweiler, R., A. Bairoch, C. H. Wu, W. C. Barker, B. Boeckmann, S. Ferro, E. Gasteiger, H. Huang, R. Lopez, M. Magrane, M. J. Martin, D. A. Natale, C. O'Donovan, N. Redaschi and L. S. Yeh (2004). "Uniprot: The Universal Protein Knowledgebase." Nucleic Acids Res **32**(Database issue): D115-119.

Aunkham, A., M. Zahn, A. Kesireddy, K. R. Pothula, A. Schulte, A. Baslé, U. Kleinekathöfer, W. Suginta and B. van den Berg (2018). "Structural Basis for Chitin Acquisition by Marine *Vibrio* Species." Nat Comms **9**(1): 220.

Bailey, D. C., E. J. Drake, T. D. Grant and A. M. Gulick (2016). "Structural and Functional Characterization of Aerobactin Synthetase Iuca from a Hypervirulent Pathotype of *Klebsiella Pneumoniae*." Biochemistry **55**(25): 3559-3570.

Bailey, K. L., S. M. Boyetchko and T. Längle (2010). "Social and Economic Drivers Shaping the Future of Biological Control: A Canadian Perspective on the Factors Affecting the Development and Use of Microbial Biopesticides." Biol Control **52**(3): 221-229.

Bain, J. (1980). "Melolonthine Beetles in Forests - Grass Grub and Other Chafers." Forest and Timber Insects in New Zealand **43**.

Ballingall, J. and R. Lattimore (2004). "Farming in New Zealand: The State of Play and Key Issues for the Backbone of the New Zealand Economy."

Bannam, T. L., X.-X. Yan, P. F. Harrison, T. Seemann, A. L. Keyburn, C. Stubenrauch, L. H. Weeramantri, J. K. Cheung, B. A. McClane, J. D. Boyce, R. J. Moore and J. I. Rood (2011). "Necrotic Enteritis-Derived *Clostridium Perfringens* Strain with Three Closely Related Independently Conjugative Toxin and Antibiotic Resistance Plasmids." MBio **2**(5): e00190-00111.

Barratt, B. I. P., V. C. Moran, F. Bigler and J. C. van Lenteren (2018). "The Status of Biological Control and Recommendations for Improving Uptake for the Future." BioControl **63**(1): 155-167.

Bascomb, S., S. P. Lapage, W. R. Willcox and M. A. Curtis (1971). "Numerical Classification of the Tribe *Klebsielleae*." Microbiology **66**(3): 279-295.

Belda, E., A. Moya, S. Bentley and F. J. Silva (2010). "Mobile Genetic Element Proliferation and Gene Inactivation Impact over the Genome Structure and Metabolic Capabilities of *Sodalis Glossinidius*, the Secondary Endosymbiont of Tsetse Flies." BMC Genom **11**(1): 449.

Benedik, M. J. and U. Strych (1998). "*Serratia Marcescens* and Its Extracellular Nuclease." FEMS Microbiol Lett **165**(1): 1-13.

Berkmen, M. and M. J. Benedik (2002). "Multi-Copy Repression of *Serratia Marcescens* Nuclease Expression by Dini." Curr Microbiol **44**(1): 44-48.

Bertelli, C., M. R. Laird, K. P. Williams, Simon Fraser University Research Computing Group, B. Y. Lau, G. Hoad, G. L. Winsor and F. S. L. Brinkman (2017). "Islandviewer 4: Expanded Prediction of Genomic Islands for Larger-Scale Datasets." Nucleic Acids Res **45**(W1): W30-W35.

Bilung, L. M., C. F. Pui, L. Su'ut and K. Apun (2018). "Evaluation of Box-Pcr and Eric-Pcr as Molecular Typing Tools for Pathogenic *Leptospira*." Dis Markers **2018**: 1351634.

Björkman, J., I. Nagaev, O. G. Berg, D. Hughes and D. I. Andersson (2000). "Effects of Environment on Compensatory Mutations to Ameliorate Costs of Antibiotic Resistance." Science **287**(5457): 1479-1482.

Black, W. A., R. Hodgson and A. McKechnie (1971). "Evaluation of Three Methods Using Deoxyribonuclease Production as a Screening Test for *Serratia Marcescens.*" Journal of Clinical Pathology **24**(4): 313-316.

Blacquière, T. and J. J. van der Steen (2017). "Three Years of Banning Neonicotinoid Insecticides Based on Sub-Lethal Effects: Can We Expect to See Effects on Bees?" Pest Manag Sci **73**(7): 1299-1304.

Blin, K., S. Shaw, K. Steinke, R. Villebro, N. Ziemert, S. Y. Lee, M. H. Medema and T. Weber (2019). "Antismash 5.0: Updates to the Secondary Metabolite Genome Mining Pipeline." Nucleic Acids Res **47**(W1): W81-W87.

Blokesch, M. and G. K. Schoolnik (2008). "The Extracellular Nuclease Dns and Its Role in Natural Transformation of *Vibrio Cholerae*." J Bacteriol **190**(21): 7232-7240.

Boetzer, M., C. V. Henkel, H. J. Jansen, D. Butler and W. Pirovano (2011). "Scaffolding Pre-Assembled Contigs Using Sspace." Bioinformatics **27**(4): 578-579.

Boetzer, M. and W. Pirovano (2012). "Toward Almost Closed Genomes with Gapfiller." Genome Biol **13**(6): R56.

Bolger, A. M., M. Lohse and B. Usadel (2014). "Trimmomatic: A Flexible Trimmer for Illumina Sequence Data." Bioinformatics **30**(15): 2114-2120.

Broudy, T. B., V. Pancholi and V. A. Fischetti (2002). "The in Vitro Interaction of *Streptococcus Pyogenes* with Human Pharyngeal Cells Induces a Phage-Encoded Extracellular Dnase." Infect and Immun **70**(6): 2805-2811.

Brown, N. F., M. E. Wickham, B. K. Coombes and B. B. Finlay (2006). "Crossing the Line: Selection and Evolution of Virulence Traits." PLOS Pathogens **2**(5): e42.

Bruce, J. B., G. A. Cooper, H. Chabas, S. A. West and A. S. Griffin (2017). "Cheating and Resistance to Cheating in Natural Populations of the Bacterium *Pseudomonas Fluorescens*." Evolution **71**(10): 2484-2495.

Brynildsrud, O., J. Bohlin, L. Scheffer and V. Eldholm (2016). "Rapid Scoring of Genes in Microbial Pan-Genome-Wide Association Studies with Scoary." Genome Biology **17**(1): 238.

Buckner, M. M. C., M. L. Ciusa and L. J. V. Piddock (2018). "Strategies to Combat Antimicrobial Resistance: Anti-Plasmid and Plasmid Curing." FEMS Microbiol Rev **42**(6): 781-804.

Buckner, M. M. C., H. T. H. Saw, R. N. Osagie, A. McNally, V. Ricci, M. E. Wand, N. Woodford, A. Ivens, M. A. Webber and L. J. V. Piddock (2018). "Clinically Relevant Plasmid-Host Interactions Indicate That Transcriptional and Not Genomic Modifications Ameliorate Fitness Costs of *Klebsiella Pneumoniae* Carbapenemase-Carrying Plasmids." MBio **9**(2): e02303-02317.

Büttner, D. and U. Bonas (2010). "Regulation and Secretion of *Xanthomonas* Virulence Factors." FEMS Microbiol Rev **34**(2): 107-133.

Cabral, C. M., A. Cherqui, A. Pereira and N. Simões (2004). "Purification and Characterization of Two Distinct Metalloproteases Secreted by the Entomopathogenic Bacterium *Photorhabdus* sp. Strain Az29." Appl Environ Microbiol **70**(7): 3831-3838.

Carattoli, A. (2009). "Resistance Plasmid Families in *Enterobacteriaceae*." Antimicrob Agents Chemother **53**(6): 2227-2238.

Carroll, A. C. and A. Wong (2018). "Plasmid Persistence: Costs, Benefits, and the Plasmid Paradox." Can J Microbiol **64**(5): 293-304.

Ceroni, F., S. Furini, T. Gorochowski, A. Boo, O. Borkowski, Y. Ladak, A. Awan, C. Gilbert, G. Stan and T. Ellis (2017). "Burden-Driven Feedback Control of Gene Expression." bioRxiv: 177030.

Chambers, R. G., K. Chatzimichael and V. Tzouvelekas (2019). "Sub-Lethal Concentrations of Neonicotinoid Insecticides at the Field Level Affect Negatively Honey Yield: Evidence from a 6-Year Survey of Greek Apiaries." PLoS One **14**(4): e0215363.

Chang, A. C. and S. N. Cohen (1978). "Construction and Characterization of Amplifiable Multicopy DNA Cloning Vehicles Derived from the P15a Cryptic Miniplasmid." J Bacteriol **134**(3): 1141-1156.

Charles, T. C. and E. W. Nester (1993). "A Chromosomally Encoded Two-Component Sensory Transduction System Is Required for Virulence of *Agrobacterium Tumefaciens*." J Bacteriol **175**(20): 6614-6625.

Chattopadhyay, P., G. Banerjee and S. Mukherjee (2017). "Recent Trends of Modern Bacterial Insecticides for Pest Control Practice in Integrated Crop Management System." 3 Biotech **7**(1): 60.

Chaudhuri, S., J. C. Bruno, F. Alonzo, 3rd, B. Xayarath, N. P. Cianciotto and N. E. Freitag (2010). "Contribution of Chitinases to *Listeria Monocytogenes* Pathogenesis." Appl Environ Microbiol **76**(21): 7302-7305.

Cheung, J. K., M. M. Awad, S. McGowan and J. I. Rood (2009). "Functional Analysis of the Virsr Phosphorelay from *Clostridium Perfringens*." PLoS One **4**(6): e5849.

Chew, S. Y., W. J. Y. Chee and L. T. L. Than (2019). "The Glyoxylate Cycle and Alternative Carbon Metabolism as Metabolic Adaptation Strategies of *Candida Glabrata*: Perspectives from *Candida Albicans* and *Saccharomyces Cerevisiae*." J Biomedical Sci **26**(1): 52-52.

Chopra, A. K., M. K. Sharma and S. Chamoli (2011). "Bioaccumulation of Organochlorine Pesticides in Aquatic System--an Overview." Environ Monit Assess **173**(1-4): 905-916.

Christensen, A. B., K. Riedel, L. Eberl, L. R. Flodgaard, S. Molin, L. Gram and M. Givskov (2003). "Quorum-Sensing-Directed Protein Expression in *Serratia Proteamaculan*s B5a." Microbiology **149**(2): 471-483.

Cimermancic, P., Marnix H. Medema, J. Claesen, K. Kurita, Laura C. Wieland Brown, K. Mavrommatis, A. Pati, Paul A. Godfrey, M. Koehrsen, J. Clardy, Bruce W. Birren, E. Takano, A. Sali, Roger G. Linington and Michael A. Fischbach (2014). "Insights into Secondary Metabolism from a Global Analysis of Prokaryotic Biosynthetic Gene Clusters." Cell **158**(2): 412-421.

Clauditz, A., A. Resch, K.-P. Wieland, A. Peschel and F. Götz (2006). "Staphyloxanthin Plays a Role in the Fitness of *Staphylococcus Aureus* and Its Ability to Cope with Oxidative Stress." Infect Immun **74**(8): 4950-4953.

Claus, H., T. A. Jackson and Z. Filip (1995). "Characterization of *Serratia Entomophila* Strains by Genomic DNA Fingerprints and Plasmid Profiles." Microbioll Res **150**(2): 159-166.

Coca-Abia, M. M. and J. Romero-Samper (2016). "Establishment of the Identity of *Costelytra Zealandica* (White 1846) (Coleoptera: Scarabeidae: Melolonthinae) a Species Commonly Known as the New Zealand Grass Grub." New Zealand Entomologist **39**(2): 129-146.

Coil, D., G. Jospin and A. E. Darling (2015). "A5-Miseq: An Updated Pipeline to Assemble Microbial Genomes from Illumina Miseq Data." Bioinformatics **31**(4): 587-589.

Comeron, J. M. (2014). "Background Selection as Baseline for Nucleotide Variation across the Drosophila Genome." PLOS Genetics **10**(6): e1004434.

Conis, E. (2010). "Debating the Health Effects of Ddt: Thomas Jukes, Charles Wurster, and the Fate of an Environmental Pollutant." Public health reports (Washington, D.C. : 1974) **125**(2): 337-342.

Coulthurst, S. J., N. R. Williamson, A. K. P. Harris, D. R. Spring and G. P. C. Salmond (2006). "Metabolic and Regulatory Engineering of *Serratia Marcescens*: Mimicking Phage-Mediated Horizontal Acquisition of Antibiotic Biosynthesis and Quorum-Sensing Capacities." Microbiology **152**(7): 1899-1911.

Cox, C. E., M. T. Brandl, M. H. de Moraes, S. Gunasekera and M. Teplitski (2018). "Production of the Plant Hormone Auxin by *Salmonella* and Its Role in the Interactions with Plants and Animals." Front Microbiol **8**(2668).

Croucher, N. J., A. J. Page, T. R. Connor, A. J. Delaney, J. A. Keane, S. D. Bentley, J. Parkhill and S. R. Harris (2014). "Rapid Phylogenetic Analysis of Large Samples of Recombinant Bacterial Whole Genome Sequences Using Gubbins." Nucleic Acids Res **43**(3): e15-e15.

Dale, S. E., M. T. Sebulsky and D. E. Heinrichs (2004). "Involvement of Sirabc in Iron-Siderophore Import in *Staphylococcus Aureus*." J Bacteriol **186**(24): 8356-8362.

Darling, A. C., B. Mau, F. R. Blattner and N. T. Perna (2004). "Mauve: Multiple Alignment of Conserved Genomic Sequence with Rearrangements." Genome Res **14**(7): 1394-1403.

Darling, A. E., B. Mau and N. T. Perna (2010). "Progressivemauve: Multiple Genome Alignment with Gene Gain, Loss and Rearrangement." PLoS One **5**(6): e11147.

Darmon, E. and D. R. F. Leach (2014). "Bacterial Genome Instability." Microbiol Mol Biol Rev : MMBR **78**(1): 1-39.

Datta, N. and R. W. Hedges (1971). "Compatibility Groups among Fi− R Factors." Nature **234**(5326): 222-223.

De Gelder, L., J. M. Ponciano, P. Joyce and E. M. Top (2007). "Stability of a Promiscuous Plasmid in Different Hosts: No Guarantee for a Long-Term Relationship." Microbiology **153**(Pt 2): 452-463.

Deatherage, D. E. and J. E. Barrick (2014). "Identification of Mutations in Laboratory Evolved Microbes from Next-Generation Sequencing Data Using Breseq." Methods in molecular biology (Clifton, N.J.) **1151**: 165-188.

del Solar, G., R. Giraldo, M. J. Ruiz-Echevarría, M. Espinosa and R. Díaz-Orejas (1998). "Replication and Control of Circular Bacterial Plasmids." Microb Mol Biol Rev : MMBR **62**(2): 434-464.

Devanga Ragupathi, N. K., D. P. Muthuirulandi Sethuvel, F. Y. Inbanathan and B. Veeraraghavan (2017). "Accurate Differentiation of *Escherichia Coli* and *Shigella* Serogroups: Challenges and Strategies." New microbes New Infect **21**: 58-62.

Dodd, S. J. (2003). Horizontal Transfer of Plasmidborne Insecticidal Toxin Genes of *Serratia* Species. Dunedin, New Zealand, University of Otago. **PhD**.

Dodd, S. J., M. R. Hurst, T. R. Glare, M. O'Callaghan and C. W. Ronson (2006). "Occurrence of Sep Insecticidal Toxin Complex Genes in *Serratia* Spp. And *Yersinia Frederiksenii*." Appl Environ Microbiol **72**(10): 6584-6592.

Dombek, P. E., L. K. Johnson, S. T. Zimmerley and M. J. Sadowsky (2000). "Use of Repetitive DNA Sequences and the Pcr to Differentiate *Escherichia Coli* Isolates from Human and Animal Sources." Appl Environ Microbiol **66**(6): 2572-2577.

Domik, D., N. Magnus and B. Piechulla (2016). "Analysis of a New Cluster of Genes Involved in the Synthesis of the Unique Volatile Organic Compound Sodorifen of *Serratia Plymuthica* 4rx13." FEMS Microbiol Lett **363**(14).

Domik, D., A. Thürmer, T. Weise, W. Brandt, R. Daniel and B. Piechulla (2016). "A Terpene Synthase Is Involved in the Synthesis of the Volatile Organic Compound Sodorifen of *Serratia Plymuthica* 4rx13." Front Microbiol **7**: 737-737.

Donati, A. J., H. I. Lee, J. H. Leveau and W. S. Chang (2013). "Effects of Indole-3-Acetic Acid on the Transcriptional Activities and Stress Tolerance of *Bradyrhizobium Japonicum*." PLoS One **8**(10): e76559.

Dowell, S. F. (2001). "Seasonal Variation in Host Susceptibility and Cycles of Certain Infectious Diseases." Emerg Infect Dis **7**(3): 369-374.

Dower, W. J., J. F. Miller and C. W. Ragsdale (1988). "High Efficiency Transformation of *E. Coli* by High Voltage Electroporation." Nucleic Acids Res **16**(13): 6127-6145.

Dulbecco, R. and M. Vogt (1954). "Plaque Formation and Isolation of Pure Lines with Poliomyelitis Viruses." J Exp Med **99**(2): 167-182.

Eddy, S. R. (2011). "Accelerated Profile Hmm Searches." PLoS Comput Biol **7**(10): e1002195.

Ehrlich, G. D., N. L. Hiller and F. Z. Hu (2008). "What Makes Pathogens Pathogenic." Genome Biol **9**(6): 225-225.

Ellwood, M. and M. Nomura (1980). "Deletion of a Ribosomal Ribonucleic Acid Operon in *Escherichia Coli*." J Bacteriol **143**(2): 1077-1080.

Environmental Protection Authority (2013). "Application for the Reassessment of a Group of Hazardous Substances." Retrieved 12 October, 2020.

Erill, I., M. Escribano, S. Campoy and J. Barbé (2003). "*In Silico* Analysis Reveals Substantial Variability in the Gene Contents of the Gamma Proteobacteria Lexa-Regulon." Bioinformatics **19**(17): 2225-2236.

Fang, W., B. Leng, Y. Xiao, K. Jin, J. Ma, Y. Fan, J. Feng, X. Yang, Y. Zhang and Y. Pei (2005). "Cloning of *Beauveria Bassiana* Chitinase Gene Bbchit1 and Its Application to Improve Fungal Strain Virulence." Appl Environ Microbiol **71**(1): 363-370.

Filloux, A. (2010). "A Variety of Bacterial Pili Involved in Horizontal Gene Transfer." J Bacteriol **192**(13): 3243-3245.

Flores-Díaz, M., L. Monturiol-Gross, C. Naylor, A. Alape-Girón and A. Flieger (2016). "Bacterial Sphingomyelinases and Phospholipases as Virulence Factors." Microbiol Mol Bio Rev **80**(3): 597.

Focareta, T. and P. A. Manning (1991). "Distinguishing between the Extracellular Dnases of *Vibrio Cholerae* and Development of a Transformation System." Mol Microbiol **5**(10): 2547-2555.

Fox, G. E., J. D. Wisotzkey and P. Jurtshuk, Jr. (1992). "How Close Is Close: 16s Rrna Sequence Identity May Not Be Sufficient to Guarantee Species Identity." Int J Syst Bacteriol **42**(1): 166-170.

Frankowski, J., M. Lorito, F. Scala, R. Schmid, G. Berg and H. Bahl (2001). "Purification and Properties of Two Chitinolytic Enzymes of *Serratia Plymuthica* Hro-C48." Arch MicrobioL **176**(6): 421-426.

Frazee, A. C., G. Pertea, A. E. Jaffe, B. Langmead, S. L. Salzberg and J. T. Leek (2015). "Ballgown Bridges the Gap between Transcriptome Assembly and Expression Analysis." Nat Biotechnol **33**(3): 243-246.

Frees, D., L. Brøndsted and H. Ingmer (2013). "Bacterial Proteases and Virulence." Subcell Biochem **66**: 161-192.

Frost, L. S., R. Leplae, A. O. Summers and A. Toussaint (2005). "Mobile Genetic Elements: The Agents of Open Source Evolution." Nat Rev Micro **3**(9): 722-732.

Fukuto, T. R. (1990). "Mechanism of Action of Organophosphorus and Carbamate Insecticides." Environ Health Perspec **87**: 245-254.

Funkhouser, J. D. and N. N. Aronson, Jr. (2007). "Chitinase Family Gh18: Evolutionary Insights from the Genomic History of a Diverse Protein Family." BMC Evol Biol **7**: 96-96.

Gabriel, M. W., G. Y. Matsui, R. Friedman and C. R. Lovell (2014). "Optimization of Multilocus Sequence Analysis for Identification of Species in the Genus *Vibrio*." Appl Environ Microbiol **80**(17): 5359-5365.

García-Fraile, P., M. Chudíčková, O. Benada, J. Pikula and M. Kolařík (2015). "*Serratia Myotis* Sp. Nov. And *Serratia Vespertilionis* Sp. Nov., Isolated from Bats Hibernating in Caves." Int J Syst Evol Microbiol **65**(Pt 1): 90-94.

Garcia-Fraile, P., C. Sproer, O. Chesneau, A. Criscuolo, E. Lang and D. Clermont (2020). "*Serratia Vespertilionis* (Garcia-Fraile Et Al. 2015) Is a Later Heterotypic Synonym of *Serratia Ficaria* (Grimont Et Al. 1981)." Int J Syst Evol Microbiol **70**(3): 1961-1962.

Gevrekci, A. (2017). "The Roles of Polyamines in Microorganisms." World J Microbiol Biotechnol **33**(11): 204.

Ghannoum, M. A. (2000). "Potential Role of Phospholipases in Virulence and Fungal Pathogenesis." Clin Microbiol Rev **13**(1): 122-143, table of contents.

Givskov, M. and S. Molin (1993). "Secretion of *Serratia Liquefaciens* Phospholipase from Escherichia Coli." Mol Microbiol **8**(2): 229-242.

Givskov, M., L. Olsen and S. Molin (1988). "Cloning and Expression in *Escherichia Coli* of the Gene for Extracellular Phospholipase A1 from *Serratia Liquefaciens*." J Bacteriol **170**(12): 5855-5862.

Glare, T. R., G. E. Corbett and T. J. Sadler (1993). "Association of a Large Plasmid with Amber Disease of the New Zealand Grass Grub, *Costelytra Zealandica*, Caused by *Serratia Entomophila* and *Serratia Proteamaculans*." J InvertebR Pathol **62**(2): 165-170.

Goforth, J. B., N. E. Walter and E. Karatan (2013). "Effects of Polyamines on *Vibrio Cholerae* Virulence Properties." PLoS One **8**(4): e60765.

Grammbitter, G. L. C., M. Schmalhofer, K. Karimi, Y. M. Shi, T. A. Schöner, N. J. Tobias, N. Morgner, M. Groll and H. B. Bode (2019). "An Uncommon Type Ii Pks Catalyzes Biosynthesis of Aryl Polyene Pigments." J Am Chem Soc **141**(42): 16615-16623.

Grant, S. G., J. Jessee, F. R. Bloom and D. Hanahan (1990). "Differential Plasmid Rescue from Transgenic Mouse Dnas into *Escherichia Coli* Methylation-Restriction Mutants." Proc Natl Acad Sci USA **87**(12): 4645-4649.

Grimont, F. a. G., A.D (2006). "The Genus *Serratia*." Prokaryotes **6**: 219-244.
Grimont, P. A. D., F. Grimont and K. Irino (1982). "Biochemical Characterization Of *Serratia Liquefaciens* Sensu Stricto, *Serratia Proteamaculans, Andserratia Grimesii* Sp. Nov." Current Microbiology **7**(2): 69-74.
Grimont, P. A. D., F. Grimont and O. Lysenko (1979). "Species and Biotype Identification of Serratia Strains Associated with Insects." Curr Microbiol **2**(3): 139-142.
Grimont, P. A. D., F. Grimont and M. P. Starr (1979). "*Serratia Ficaria* Sp. Nov., a Bacterial Species Associated with Smyrna Figs and the Fig Wasp *Blastophaga Psenes*." Curr Microbiol **2**(5): 277-282.
Grimont, P. A. D., F. Grimont and M. P. Starr (1981). "*Serratia* Species Isolated from Plants." Curr Microbiol **5**(5): 317-322.
Grimont, P. A. D., T. A. Jackson, E. Ageron and M. J. Noonan (1988). "*Serratia Entomophila* Sp. Nov. Associated with Amber Disease in the New Zealand Grass Grub *Costelytra Zealandica.*" Int J Sys Bacteriol **38**(1): 1-6.
Grkovic, S., T. R. Glare, T. A. Jackson and G. E. Corbett (1995). "Genes Essential for Amber Disease in Grass Grubs Are Located on the Large Plasmid Found in *Serratia Entomophila* and *Serratia Proteamaculans*." Appl Environ Microbiol **61**(6): 2218-2223.
Grkovic, S. and H. K. Mahanty (1996). "Investigation of a Phage Resistant *Serratia Entomophila* Strain (Bc4b), Establishment of Generalised Transduction and Construction of *S. Entomophila* Reca Mutants." Mol Gen Genet **250**(3): 323-328.
Gyles, C. and P. Boerlin (2014). "Horizontally Transferred Genetic Elements and Their Role in Pathogenesis of Bacterial Disease." Vet Pathol **51**(2): 328-340.
Hahn, M. W., U. Koll, J. Jezberová and A. Camacho (2015). "Global Phylogeography of Pelagic *Polynucleobacter* Bacteria: Restricted Geographic Distribution of Subgroups, Isolation by Distance and Influence of Climate." Environ Microbiol **17**(3): 829-840.
Hakbijl, T. (2002). "The Traditional, Historical and Prehistoric Use of Ashes as an Insecticide, with an Experimental Study on the Insecticidal Efficacy of Washed Ash." Environl Archaeol **7**(1): 13-22.
Hall, J. P. J., A. J. Wood, E. Harrison and M. A. Brockhurst (2016). "Source–Sink Plasmid Transfer Dynamics Maintain Gene Mobility in Soil Bacterial Communities."PNAS **113**(29): 8260-8265.
Han, N., Y. Qiang and W. Zhang (2016). "Anitools Web: A Web Tool for Fast Genome Comparison within Multiple Bacterial Strains." Database (Oxford) **2016**.
Hanson, C. A., J. A. Fuhrman, M. C. Horner-Devine and J. B. Martiny (2012). "Beyond Biogeographic Patterns: Processes Shaping the Microbial Landscape." Nat Rev Microbiol **10**(7): 497-506.
Harrison, E. and M. A. Brockhurst (2012). "Plasmid-Mediated Horizontal Gene Transfer Is a Coevolutionary Process." Trends Microbiol **20**(6): 262-267.
Hayek, N. (2013). "Lateral Transfer and Gc Content of Bacterial Resistant Genes." Front Microbiol **4**: 41-41.
Hayes, F. (2003). "Toxins-Antitoxins: Plasmid Maintenance, Programmed Cell Death, and Cell Cycle Arrest." Science **301**(5639): 1496-1499.
He, J., R. L. Baldini, E. Déziel, M. Saucier, Q. Zhang, N. T. Liberati, D. Lee, J. Urbach, H. M. Goodman and L. G. Rahme (2004). "The Broad Host Range Pathogen *Pseudomonas Aeruginosa* Strain Pa14 Carries Two Pathogenicity Islands Harboring Plant and Animal Virulence Genes." Proc Natl Acad Sci USA **101**(8): 2530-2535.
Hersch, S. and W. Navarre (2019). "*Salmonella* Itar Responds to Itaconate in Macrophage." bioRxiv: 648865.
Hershberg, R. (2015). "Mutation--the Engine of Evolution: Studying Mutation and Its Role in the Evolution of Bacteria." Cold Spring Harb Perspect Biol **7**(9): a018077-a018077.
Hirschmann, M., F. Grundmann and H. B. Bode (2017). "Identification and Occurrence of the Hydroxamate Siderophores Aerobactin, Putrebactin, Avaroferrin and Ochrobactin C as Virulence Factors from Entomopathogenic Bacteria." Environ Microbiol **19**(10): 4080-4090.
Ho Sui, S. J., A. Fedynak, W. W. L. Hsiao, M. G. I. Langille and F. S. L. Brinkman (2009). "The Association of Virulence Factors with Genomic Islands." PLoS One **4**(12): e8094-e8094.

Hooper, S. D., K. Mavromatis and N. C. Kyrpides (2009). "Microbial Co-Habitation and Lateral Gene Transfer: What Transposases Can Tell Us." Genome Biol **10**(4): R45.
Hossain, A. H., A. Ter Beek and P. J. Punt (2019). "Itaconic Acid Degradation in *Aspergillus Niger*: The Role of Unexpected Bioconversion Pathways." Fungal Biology and Biotechnology **6**(1): 1.
Hsiao, W., I. Wan, S. Jones and F. Brinkman (2003). "Islandpath: Aiding Detection of Genomic Islands in Prokaryotes." Bioinformatics (Oxford, England) **19**: 418-420.
Hsu, S. C. and J. L. Lockwood (1975). "Powdered Chitin Agar as a Selective Medium for Enumeration of Actinomycetes in Water and Soil." Appl Microbiol **29**(3): 422-426.
Hunt, M., N. D. Silva, T. D. Otto, J. Parkhill, J. A. Keane and S. R. Harris (2015). "Circlator: Automated Circularization of Genome Assemblies Using Long Sequencing Reads." Genome Biol **16**: 294.
Hurst, M. R., A. K. Beattie, S. A. Jones, P. C. Hsu, J. Calder and C. van Koten (2015). "Temperature-Dependent *Galleria Mellonella* Mortality as a Result of *Yersinia Entomophaga* Infection." Appl Environ Microbiol **81**(18): 6404-6414.
Hurst, M. R., S. A. Becher and M. O'Callaghan (2011). "Nucleotide Sequence of the *Serratia Entomophila* Plasmid Padap and the *Serratia Proteamaculans* Pu143 Plasmid Virulence Associated Region." Plasmid **65**(1): 32-41.
Hurst, M. R., S. A. Becher, S. D. Young, T. L. Nelson and T. R. Glare (2011). "*Yersinia Entomophaga* Sp. Nov., Isolated from the New Zealand Grass Grub *Costelytra Zealandica*." Int J Syst Evol Microbiol **61**(Pt 4): 844-849.
Hurst, M. R., S. M. Jones, B. Tan and T. A. Jackson (2007). "Induced Expression of the *Serratia Entomophila* Sep Proteins Shows Activity Towards the Larvae of the New Zealand Grass Grub *Costelytra Zealandica*." FEMS Microbiol Lett **275**(1): 160-167.
Hurst, M. R. H. (2016). Non- Spore Forming Bacterial Entomopathogens -Their Toxins, Hosts and the Environment: Why Be a Pathogen. The Rasputin Effect: When Commensals and Symbionts Become Parasitic. C. J. Hurst, Springer. **3:** 84-100.
Hurst, M. R. H., A. Beattie, S. A. Jones, A. Laugraud, C. van Koten and L. Harper (2018). "*Serratia Proteamaculans* Strain Agr96x Encodes an Antifeeding Prophage (Tailocin) with Activity against Grass Grub (*Costelytra Giveni*) and Manuka Beetle (*Pyronota* Species) Larvae." Appl Environ Microbiol **84**(10).
Hurst, M. R. H., S. A. Becher and M. O'Callaghan (2011). "Nucleotide Sequence of the *Serratia Entomophila* Plasmid Padap and the *Serratia Proteamaculans* Pu143 Plasmid Virulence Associated Region." Plasmid **65**(1): 32-41.
Hyatt, D., G. L. Chen, P. F. Locascio, M. L. Land, F. W. Larimer and L. J. Hauser (2010). "Prodigal: Prokaryotic Gene Recognition and Translation Initiation Site Identification." BMC Bioinformatics **11**: 119.
Isaeva, A. S., E. E. Kulikov, K. K. Tarasyan and A. V. Letarov (2010). "A Novel High-Resolving Method for Genomic Pcr-Fingerprinting of Enterobacteria." Acta Naturae **2**(1): 82-88.
Jackson, T. (1990). "Biological Control of Grass Grub in Canterbury." Proc. N.Z. Grassl. Assoc. **52**.
Jackson, T., T. Glare, M. O'Callaghan (1991). "Pathotypic Boundaries for *Serratia* Spp. Causing Amber Disease in the New Zealand Grass Grub, *Costelytra Zealandica*." Proceedings of the 3rd European Meeting of Microbial Control of Pests **61**: 123-130.
Jackson, T. A. (1999). "Factors in the Success and Failure of Microbial Controlagents for Soil Dwelling Pests." Int Pest Manag Rev **4**(4): 281-285.
Jackson, T. A., D. G. Boucias and J. O. Thaler (2001). "Pathobiology of Amber Disease, Caused by *Serratia* Spp., in the New Zealand Grass Grub, *Costelytra Zealandica.*" J Invertebr Pathol **78**(4): 232-243.
Jackson, T. A., A. M. Huger and T. R. Glare (1993). "Pathology of Amber Disease in the New Zealand Grass Grub *Costelytra Zealandica* (Coleoptera: Scarabaeidae)." J Invert Path **61**(2): 123-130.
Javidpour, P., S. Deutsch, V. K. Mutalik, N. J. Hillson, C. J. Petzold, J. D. Keasling and H. R. Beller (2016). "Investigation of Proposed Ladderane Biosynthetic Genes from Anammox Bacteria by Heterologous Expression in *E. Coli.*" PLoS One **11**(3): e0151087.
Jiménez, J. I., Á. Canales, J. Jiménez-Barbero, K. Ginalski, L. Rychlewski, J. L. García and E. Díaz (2008). "Deciphering the Genetic Determinants for Aerobic Nicotinic Acid Degradation: The Nic Cluster from *Pseudomonas Putida* Kt2440." PNAS **105**(32): 11329-11334.

Jin, S., Y. Chen, G. E. Christie and M. J. Benedik (1996). "Regulation of the *Serratia Marcescens* Extracellular Nuclease: Positive Control by a Homolog of P2 Ogr Encoded by a Cryptic Prophage." J Mol Biol **256**(2): 264-278.

Johnson, V. (2004). "Establishment of *Serratia Entomophila* after Application of a New Formulation for Grass Grub Control." New Zealand Plant Protection **57**: 310-313.

Johnson, V., J. Pearson and T. Jackson (2001). "Formulation of *Serratia Entomophila* for Biological Control of Grass Grub." New Zealand Plant Protection **54**: 125-127.

Jones, J. D., K. L. Grady, T. V. Suslow and J. R. Bedbrook (1986). "Isolation and Characterization of Genes Encoding Two Chitinase Enzymes from *Serratia Marcescens*." The EMBO journal **5**(3): 467-473.

Juhas, M., J. R. van der Meer, M. Gaillard, R. M. Harding, D. W. Hood and D. W. Crook (2009). "Genomic Islands: Tools of Bacterial Horizontal Gene Transfer and Evolution." FEMS Microbiol Rev **33**(2): 376-393.

Kado, C. I. and S. T. Liu (1981). "Rapid Procedure for Detection and Isolation of Large and Small Plasmids." J Bacteriol **145**(3): 1365-1373.

Kalayu, G. (2019). "Phosphate Solubilizing Microorganisms: Promising Approach as Biofertilizers." Int J Agron **2019**: 4917256.

Kämpfer, P. and S. P. Glaeser (2015). "*Serratia Glossinae* Geiger Et Al. 2010 Is a Later Synonym of Serratia Fonticola Gavini Et Al. 1979." INTJ Syst Evol Microbiol **65**(Pt_5): 1406-1408.

Kämpfer, P. and S. P. Glaeser (2016). "*Serratia Aquatilis* Sp. Nov., Isolated from Drinking Water Systems." Int J Syst EvolMicrobiol **66**(1): 407-413.

Kasprzewska, A. (2003). "Plant Chitinases--Regulation and Function." Cell Mol Biol Lett **8**(3): 809-824.

Kawase, T., A. Saito, T. Sato, R. Kanai, T. Fujii, N. Nikaidou, K. Miyashita and T. Watanabe (2004). "Distribution and Phylogenetic Analysis of Family 19 Chitinases in *Actinobacteria*." Appl Environ Microbiol **70**(2): 1135-1144.

Kelley, L. A., S. Mezulis, C. M. Yates, M. N. Wass and M. J. E. Sternberg (2015). "The Phyre2 Web Portal for Protein Modeling, Prediction and Analysis." Nat prot **10**(6): 845-858.

Khilyas, I. V., T. V. Shirshikova, L. E. Matrosova, A. V. Sorokina, M. R. Sharipova and L. M. Bogomolnaya (2016). "Production of Siderophores by *Serratia Marcescens* and the Role of Macab Efflux Pump in Siderophores Secretion." BioNanoScience **6**(4): 480-482.

Kilian, M., D. R. Riley, A. Jensen, H. Brüggemann and H. Tettelin (2014). "Parallel Evolution of *Streptococcus Pneumoniae* and *Streptococcus Mitis* to Pathogenic and Mutualistic Lifestyles." MBio **5**(4): e01490.

Kim, M., H. S. Oh, S. C. Park and J. Chun (2014). "Towards a Taxonomic Coherence between Average Nucleotide Identity and 16s Rrna Gene Sequence Similarity for Species Demarcation of Prokaryotes." Int J Syst Evol Microbiol **64**(Pt 2): 346-351.

Kirn, T. J., B. A. Jude and R. K. Taylor (2005). "A Colonization Factor Links *Vibrio Cholerae* Environmental Survival and Human Infection." Nature **438**(7069): 863-866.

Klappenbach, J. A., J. M. Dunbar and T. M. Schmidt (2000). "Rrna Operon Copy Number Reflects Ecological Strategies of Bacteria." Appl Environ Microbiol **66**(4): 1328-1333.

Köhl, J., R. Kolnaar and W. J. Ravensberg (2019). "Mode of Action of Microbial Biological Control Agents against Plant Diseases: Relevance Beyond Efficacy." Front Plant Sci **10**(845).

Kolar, S. L., J. A. Ibarra, F. E. Rivera, J. M. Mootz, J. E. Davenport, S. M. Stevens, A. R. Horswill and L. N. Shaw (2013). "Extracellular Proteases Are Key Mediators of *Staphylococcus Aureus* Virulence Via the Global Modulation of Virulence-Determinant Stability." Microbiologyopen **2**(1): 18-34.

Kolbe, D. L. and S. R. Eddy (2011). "Fast Filtering for Rna Homology Search." Bioinformatics **27**(22): 3102-3109.

Koren, S., B. P. Walenz, K. Berlin, J. R. Miller, N. H. Bergman and A. M. Phillippy (2017). "Canu: Scalable and Accurate Long-Read Assembly Via Adaptive K-Mer Weighting and Repeat Separation." Genome Res **27**(5): 722-736.

Kumar, S., G. Stecher and K. Tamura (2016). "Mega7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets." Mol Biol Evol **33**(7): 1870-1874.

Kurokawa, M., S. Seno, H. Matsuda and B.-W. Ying (2016). "Correlation between Genome Reduction and Bacterial Growth." DNA Research **23**(6): 517-525.
Kwong, T. C. (2002). "Organophosphate Pesticides: Biochemistry and Clinical Toxicology." Ther Drug Monit **24**(1): 144-149.
Lagesen, K., P. Hallin, E. A. Rodland, H. H. Staerfeldt, T. Rognes and D. W. Ussery (2007). "Rnammer: Consistent and Rapid Annotation of Ribosomal Rna Genes." Nucleic Acids Res **35**(9): 3100-3108.
Lam, K. L., S. Lin, C. Liu, X. Wu, S. Tang, H. S. Kwan and P. C. Cheung (2018). "Low-Cost Method Generating *in Situ* Anaerobic Conditions on a 96-Well Plate for Microbial Fermentation in Food Research." J Agric Food Chem **66**(44): 11839-11845.
Landsberg, M. J., S. A. Jones, R. Rothnagel, J. N. Busby, S. D. Marshall, R. M. Simpson, J. S. Lott, B. Hankamer and M. R. Hurst (2011). "3d Structure of the *Yersinia Entomophaga* Toxin Complex and Implications for Insecticidal Activity." Proc Natl Acad Sci USA **108**(51): 20544-20549.
Langille, M. G., W. W. Hsiao and F. S. Brinkman (2008). "Evaluation of Genomic Island Predictors Using a Comparative Genomics Approach." BMC Bioinformatics **9**: 329.
Laslett, D. and B. Canback (2004). "Aragorn, a Program to Detect Trna Genes and Tmrna Genes in Nucleotide Sequences." Nucleic Acids Res **32**(1): 11-16.
Lawrence, J. G. and H. Ochman (1997). "Amelioration of Bacterial Genomes: Rates of Change and Exchange." J Mol Evol **44**(4): 383-397.
Lechner, M., S. Findeiß, L. Steiner, M. Marz, P. F. Stadler and S. J. Prohaska (2011). "Proteinortho: Detection of (Co-)Orthologs in Large-Scale Analysis." BMC Bioinformatics **12**(1): 124.
Lefort, M.-C., S. Boyer, J. Vereijssen, R. Sprague, T. R. Glare and S. P. Worner (2015). "Preference of a Native Beetle for "Exoticism," Characteristics That Contribute to Invasive Success of *Costelytra Zealandica* (Scarabaeidae: Melolonthinae)." PeerJ **3**: e1454.
Lefort, M. C., B. I. P. Barratt, J. W. M. Marris and S. Boyer (2013). "Combining Molecular and Morphological Approaches to Differentiate the Pest *Costelytra Zealandica* (White) (Coleoptera: Scarabeidae: Melolonthinae) from the Non-Pest *Costelytra Brunneum* (Broun) at the Larval Stage." N Z Entomol **36**(1): 15-21.
Leidich, S. D., A. S. Ibrahim, Y. Fu, A. Koul, C. Jessup, J. Vitullo, W. Fonzi, F. Mirbod, S. Nakashima, Y. Nozawa and M. A. Ghannoum (1998). "Cloning and Disruption of Caplb1, a Phospholipase B Gene Involved in the Pathogenicity of *Candida Albicans*." J Biol Chem **273**(40): 26078-26086.
Li, J., Y. Yao, H. H. Xu, L. Hao, Z. Deng, K. Rajakumar and H. Y. Ou (2015). "Secret6: A Web-Based Resource for Type Vi Secretion Systems Found in Bacteria." Environ Microbiol **17**(7): 2196-2202.
Li, X., E. M. Top, Y. Wang, C. J. Brown, F. Yao, S. Yang, Y. Jiang and H. Li (2015). "The Broad-Host-Range Plasmid Psfa231 Isolated from Petroleum-Contaminated Sediment Represents a New Member of the Proma Plasmid Family." Front Microbiol **5**(777).
Little, J. W. (1984). "Autodigestion of Lexa and Phage Lambda Repressors." Proc Natl Acad Sci U S A **81**(5): 1375-1379.
López-Hermoso, C., R. R. de la Haba, C. Sánchez-Porro, R. T. Papke and A. Ventosa (2017). "Assessment of Multilocus Sequence Analysis as a Valuable Tool for the Classification of the Genus *Salinivibrio.*" Front Microbiol **8**: 1107-1107.
Lysenko, O. (1976). "Chitinase of *Serratia Marcescens* and Its Toxicity to Insects." J Invert Path **27**: 385-286.
Magnus, N., T. Weise and B. Piechulla (2017). "Carbon Catabolite Repression Regulates the Production of the Unique Volatile Sodorifen of *Serratia Plymuthica* 4rx13." Frontiers in Microbiology **8**(2522).
Mahlen, S. D. (2011). "Serratia Infections: From Military Experiments to Current Practice." Clin Microbiol Rev **24**(4): 755-791.
Manzano-Marín, A. and A. Latorre (2016). "Snapshots of a Shrinking Partner: Genome Reduction in *Serratia Symbiotica*." Scientific Reports **6**(1): 32590.
Marshall, N. C., B. B. Finlay and C. M. Overall (2017). "Sharpening Host Defenses During Infection: Proteases Cut to the Chase." Mol Cell Proteomics **16**(4 suppl 1): S161-S171.

Martin, B., O. Humbert, M. Camara, E. Guenzi, J. Walker, T. Mitchell, P. Andrew, M. Prudhomme, G. Alloing, R. Hakenbeck and et al. (1992). "A Highly Conserved Repeated DNA Element Located in the Chromosome of *Streptococcus Pneumoniae*." Nucleic Acids Res **20**(13): 3479-3483.

Maslowska, K. H., K. Makiela-Dzbenska and I. J. Fijalkowska (2019). "The Sos System: A Complex and Tightly Regulated Response to DNA Damage." Environ Mol Mutagen **60**(4): 368-384.

Matelska, D., M. Kurkowska, E. Purta, J. M. Bujnicki and S. Dunin-Horkawicz (2016). "Loss of Conserved Noncoding Rnas in Genomes of Bacterial Endosymbionts." Genome biol Evol **8**(2): 426-438.

Maurelli, A. T., R. E. Fernández, C. A. Bloch, C. K. Rode and A. Fasano (1998). "“Black Holes” and Bacterial Pathogenicity: A Large Genomic Deletion That Enhances the Virulence of *Shigella* Spp. And Enteroinvasive *Escherichia Coli*." PNAS **95**(7): 3943-3948.

Mc Carlie, S. J., J. E. Hellmuth, J. Newman, C. E. Boucher and R. R. Bragg (2020). "Genome Sequence of Resistant *Serratia* Sp. Strain Hri, Isolated from a Bottle of Didecyldimethylammonium Chloride-Based Disinfectant." Microbiol Res Announc **9**(18): e00095-00020.

McLennan, J. A. and B. W. H. MacMillan (1983). "Predation by the Rook, *Corvus Frugilegus* L.,on Larvae of the Grass Grub, *Costelytra Zealandica* (White),in Hawke’s Bay, New Zealand." New Zealand Journal of Agricultural Research **26**(1): 139-145.

Mee, M. T., J. J. Collins, G. M. Church and H. H. Wang (2014). "Syntrophic Exchange in Synthetic Microbial Communities." PNAS **111**(20): E2149-E2156.

Merrikh, C. N. and H. Merrikh (2018). "Gene Inversion Potentiates Bacterial Evolvability and Virulence." Nat CommS **9**(1): 4662.

Merzendorfer, H. and L. Zimoch (2003). "Chitin Metabolism in Insects: Structure, Function and Regulation of Chitin Synthases and Chitinases." J Exp Biol **206**(Pt 24): 4393-4412.

Metcalf, W. W., W. Jiang and B. L. Wanner (1994). "Use of the Rep Technique for Allele Replacement to Construct New *Escherichia Coli* Hosts for Maintenance of R6k Gamma Origin Plasmids at Different Copy Numbers." Gene **138**(1-2): 1-7.

Miln, A. J. (1978). "Protozoan Parasites of *Costelytra Zealandica* (Coleoptera: Scarabaeidae) in New Zealand." N Z Entomol **6**(4): 392-399.

Ministry for the Environment and Stats NZ (2018). New Zealand’s Environmental Reporting Series: Our Land 2018. New Zealand.

Miyoshi, S.-I. (2013). "Extracellular Proteolytic Enzymes Produced by Human Pathogenic *Vibrio Species*." Front Microbiol **4**(339).

Modi, R. I. and J. Adams (1991). "Coevolution in Bacterial-Plasmid Populations." Evolution **45**(3): 656-667.

Moran, N. A. (2002). "Microbial Minimalism." Cell **108**(5): 583-586.

Morimoto, T., R. Kadoya, K. Endo, M. Tohata, K. Sawada, S. Liu, T. Ozawa, T. Kodama, H. Kakeshita, Y. Kageyama, K. Manabe, S. Kanaya, K. Ara, K. Ozaki and N. Ogasawara (2008). "Enhanced Recombinant Protein Productivity by Genome Reduction in *Bacillus Subtilis*." DNA Research **15**(2): 73-81.

Mortier-Barrière, I., M. Velten, P. Dupaigne, N. Mirouze, O. Piétrement, S. McGovern, G. Fichant, B. Martin, P. Noirot, E. Le Cam, P. Polard and J.-P. Claverys (2007). "A Key Presynaptic Role in Transformation for a Widespread Bacterial Protein: Dpra Conveys Incoming Ssdna to Reca." Cell **130**(5): 824-836.

Müller, A. U., F. Imkamp and E. Weber-Ban (2018). "The Mycobacterial Lexa/Reca-Independent DNA Damage Response Is Controlled by Pafbc and the Pup-Proteasome System." Cell Rep **23**(12): 3551-3564.

Nhieu, G. T. and P. J. Sansonetti (1999). "Mechanism of *Shigella* Entry into Epithelial Cells." Curr Opin Microbiol **2**(1): 51-55.

Nicolopoulou-Stamati, P., S. Maipas, C. Kotampasi, P. Stamatis and L. Hens (2016). "Chemical Pesticides and Human Health: The Urgent Need for a New Concept in Agriculture." Front Pub Health **4**(148).

Nuñez-Valdez, M. E., M. A. Calderón, E. Aranda, L. Hernández, R. M. Ramírez-Gama, L. Lina, Z. Rodríguez-Segura, M. d. C. Gutiérrez and F. J. Villalobos (2008). "Identification of a Putative

Mexican Strain of *Serratia Entomophila* Pathogenic against Root-Damaging Larvae of Scarabaeidae (Coleoptera)." Appl Environ Microbiol **74**(3): 802.
O'Callaghan, M. (1989). "The Soil Ecology of *Serratia Entomophila*, a Bacterial Pathogen of the New Zealand Grass Grub, *Costelytra Zealandica* (White)."
O'Callaghan, M. and T. A. Jackson (1993). "Isolation and Enumeration of *Serratia Entomophila*—a Bacterial Pathogen of the New Zealand Grass Grub, *Costelytra Zealandica.*" J App Bacteriol **75**(4): 307-314.
Ochman, H., J. G. Lawrence and E. A. Groisman (2000). "Lateral Gene Transfer and the Nature of Bacterial Innovation." Nature **405**(6784): 299-304.
Ohtani, K., Y. Yuan, S. Hassan, R. Wang, Y. Wang and T. Shimizu (2009). "Virulence Gene Regulation by the Agr System in *Clostridium Perfringens*." J Bacteriol **191**(12): 3919.
Page, A. J., C. A. Cummins, M. Hunt, V. K. Wong, S. Reuter, M. T. Holden, M. Fookes, D. Falush, J. A. Keane and J. Parkhill (2015). "Roary: Rapid Large-Scale Prokaryote Pan Genome Analysis." Bioinformatics **31**(22): 3691-3693.
Page, R. and W. Peti (2016). "Toxin-Antitoxin Systems in Bacterial Growth Arrest and Persistence." Nat Chem Biol **12**(4): 208-214.
Palma, L. (2017). "*Bacillus Thuringiensis*-Based Biopesticides, Are They as Effective as They Should Be?" Rev Argent Microbiol **49**(1): 119.
Park, M., E. Do and W. H. Jung (2013). "Lipolytic Enzymes Involved in the Virulence of Human Pathogenic Fungi." Mycobiology **41**(2): 67-72.
Park, S.-C., K. Lee, Y. O. Kim, S. Won and J. Chun (2019). "Large-Scale Genomics Reveals the Genetic Characteristics of Seven Species and Importance of Phylogenetic Distance for Estimating Pan-Genome Size." Front Microbiol **10**(834).
Parkins, M. D., H. Ceri and D. G. Storey (2001). "*Pseudomonas Aeruginosa Gaca*, a Factor in Multihost Virulence, Is Also Essential for Biofilm Formation." Mol Microbiol **40**(5): 1215-1226.
Parsons, J. A., T. L. Bannam, R. J. Devenish and J. I. Rood (2007). "Tcpa, an Ftsk/Spoiiie Homolog, Is Essential for Transfer of the Conjugative Plasmid Pcw3 in *Clostridium Perfringens*." J Bacteriol **189**(21): 7782.
Pearson, J. F. J., T.A (1995). "Quality Control Management of the Grass Grub Microbial Control Product, Invade®." Proceedings Agronomy Society of New Zealand **25**.
Pertea, M., G. M. Pertea, C. M. Antonescu, T.-C. Chang, J. T. Mendell and S. L. Salzberg (2015). "Stringtie Enables Improved Reconstruction of a Transcriptome from Rna-Seq Reads." Nat Biotechnol **33**(3): 290-295.
Petersen, L. M. and L. S. Tisa (2014). "Molecular Characterization of Protease Activity in *Serratia* Sp. Strain Scbi and Its Importance in Cytotoxicity and Virulence." J Bacteriol **196**(22): 3923-3936.
Petersen, T. N., S. Brunak, G. von Heijne and H. Nielsen (2011). "Signalp 4.0: Discriminating Signal Peptides from Transmembrane Regions." Nat Methods **8**(10): 785-786.
Pikovskaya, R. I. (1948). "Mobilization of Phosphorus in Soil in Connection with Vital Activity of Some Microbial Species." Mikrobiologiya **17**: 362-370.
Popa, O., E. Hazkani-Covo, G. Landan, W. Martin and T. Dagan (2011). "Directed Networks Reveal Genomic Barriers and DNA Repair Bypasses to Lateral Gene Transfer among Prokaryotes." Genome Res **21**(4): 599-609.
Poplawsky, A. R., S. C. Urban and W. Chun (2000). "Biological Role of Xanthomonadin Pigments in *Xanthomonas Campestris Pv. Campestris.*" Appl Environ Microbiol **66**(12): 5123-5127.
Poretsky, R., L. M. Rodriguez-R, C. Luo, D. Tsementzi and K. T. Konstantinidis (2014). "Strengths and Limitations of 16s Rrna Gene Amplicon Sequencing in Revealing Temporal Microbial Community Dynamics." PLoS One **9**(4): e93827.
Pujol, C., J. P. Grabenstein, R. D. Perry and J. B. Bliska (2005). "Replication of *Yersinia Pestis* in Interferon Γ-Activated Macrophages Requires *Ripa*, a Gene Encoded in the Pigmentation Locus." Proc Natl Acad Sci USA **102**(36): 12909-12914.
Qian, Y., H.-H. Huang, J. I. Jiménez and D. Del Vecchio (2017). "Resource Competition Shapes the Response of Genetic Circuits." ACS Synthetic Biology **6**(7): 1263-1272.

Qin, J., Z. Tong, Y. Zhan, C. Buisson, F. Song, K. He, C. Nielsen-LeRoux and S. Guo (2020). "A Bacillus Thuringiensis Chitin-Binding Protein Is Involved in Insect Peritrophic Matrix Adhesion and Takes Part in the Infection Process." Toxins **12**(4): 252.

Qin, Q.-L., Y. Li, L.-L. Sun, Z.-B. Wang, S. Wang, X.-L. Chen, A. Oren and Y.-Z. Zhang (2019). "Trophic Specialization Results in Genomic Reduction in Free-Living Marine *Idiomarina* Bacteria." MBio **10**(1): e02545-02518.

Ramos-Morales, F. (2012). "Impact of *Salmonella Enterica* Type Iii Secretion System Effectors on the Eukaryotic Host Cell." ISRN Cell Biology **2012**: 36.

Ramsay, J. P., M. F. Hynes, J. T. Sullivan and C. W. Ronson (2017). Symbiosis Islands. Module in Life Sciences, Elsevier.

Rasamiravaka, T., Q. Labtani, P. Duez and M. El Jaziri (2015). "The Formation of Biofilms by *Pseudomonas Aeruginosa*: A Review of the Natural and Synthetic Compounds Interfering with Control Mechanisms." BioMed Res Int **2015**: 759348-759348.

Rawlings, N. D., A. J. Barrett and A. Bateman (2010). "Merops: The Peptidase Database." Nucleic Acids Res **38**(Database issue): D227-233.

Reen, F. J., J. M. Haynes, M. J. Mooij and F. O'Gara (2013). "A Non-Classical Lysr-Type Transcriptional Regulator Pa2206 Is Required for an Effective Oxidative Stress Response in *Pseudomonas Aeruginosa."* PLoS One **8**(1): e54479.

Rice, S. A., K. S. Koh, S. Y. Queck, M. Labbate, K. W. Lam and S. Kjelleberg (2005). "Biofilm Formation and Sloughing in *Serratia Marcescens* Are Controlled by Quorum Sensing and Nutrient Cues." J Bacteriol **187**(10): 3477-3485.

Richards, N. K., T. R. Glare and D. C. A. Hall (1997). "Genetic Variation in Grass Grub, *Costelytra Zealandica*, from Several Regions." New Zealand Plant Protection **50**.

Riedel, T., M. Rohlfs, I. Buchholz, I. Wagner-Dobler and M. Reck (2013). "Complete Sequence of the Suicide Vector Pjp5603." Plasmid **69**(1): 104-107.

Rodriguez-Kabana, R., G. Godoy, G. Morgan-Jones and R. A. Shelby (1983). "The Determination of Soil Chitinase Activity: Conditions for Assay and Ecological Studies." Plant and Soil **75**(1): 95-106.

Rolle, R., A. Ejiofor and T. Johnson (2005). "Determination of the Plasmid Size and Location of D-Endotoxin Genes of *Bacillus Thuringiensis* by Pulse Field Gel Electrophoresis." African J of Biotechnol **4**: 580-585.

Rønn, R., X. Hao, F. Lüthje, N. A. German, X. Li, F. Huang, J. Kisaka, D. Huffman, H. A. Alwathnani, Y.-G. Zhu and C. Rensing (2017). "Bacterial Survival in Dictyostelium." Bio-protocol **7**(13): e2376.

Roth, J., N. Benson, T. Galitski, K. Haack, J. G. Lawrence and L. Miesel (1996). "Rearrangements of the Bacterial Chromosome: Formation and Applications." Cell Mol Biol **2**: 2256-2276.

Ruiu, L. (2013). "*Brevibacillus Laterosporus*, a Pathogen of Invertebrates and a Broad-Spectrum Antimicrobial Species." Insects **4**(3): 476-492.

Russo, T. A., R. Olson, U. MacDonald, J. Beanan and B. A. Davidson (2015). "Aerobactin, but Not Yersiniabactin, Salmochelin, or Enterobactin, Enables the Growth/Survival of Hypervirulent (Hypermucoviscous) *Klebsiella Pneumoniae*." Infection and Immunity **83**(8): 3325-3333.

Russo, T. A., R. Olson, U. MacDonald, D. Metzger, L. M. Maltese, E. J. Drake and A. M. Gulick (2014). "Aerobactin Mediates Virulence and Accounts for Increased Siderophore Production under Iron-Limiting Conditions by Hypervirulent (Hypermucoviscous) *Klebsiella Pneumoniae*." Infection and Immunity **82**(6): 2356-2367.

Russo, T. A., A. S. Shon, J. M. Beanan, R. Olson, U. MacDonald, A. O. Pomakov and M. P. Visitacion (2011). "Hypervirulent K. Pneumoniae Secretes More and More Active Iron-Acquisition Molecules Than "Classical" *K. Pneumoniae* Thereby Enhancing Its Virulence." PLoS One **6**(10): e26734.

Rutherford, S. T. and B. L. Bassler (2012). "Bacterial Quorum Sensing: Its Role in Virulence and Possibilities for Its Control." Cold Spring Harb Perspect Med **2**(11): a012427.

Rybakova, D., M. Radjainia, A. Turner, A. Sen, A. K. Mitra and M. R. Hurst (2013). "Role of Antifeeding Prophage (Afp) Protein Afp16 in Terminating the Length of the Afp Tailocin and Stabilizing Its Sheath." Mol Microbiol **89**(4): 702-714.

Saavedra De Bast, M., N. Mine and L. Van Melderen (2008). "Chromosomal Toxin-Antitoxin Systems May Act as Antiaddiction Modules." J Bacteriol **190**(13): 4603-4609.

San Millan, A., J. A. Escudero, D. R. Gifford, D. Mazel and R. C. MacLean (2016). "Multicopy Plasmids Potentiate the Evolution of Antibiotic Resistance in Bacteria." Nat Ecol & Evol **1**: 0010.

San Millan, A., M. Toll-Riera, Q. Qi, A. Betts, R. J. Hopkinson, J. McCullagh and R. C. MacLean (2018). "Integrative Analysis of Fitness and Metabolic Effects of Plasmids in *Pseudomonas Aeruginosa* Pao1." The ISME Journal **12**(12): 3014-3024.

Sandoval-Herrera, N., F. Mena, M. Espinoza and A. Romero (2019). "Neurotoxicity of Organophosphate Pesticides Could Reduce the Ability of Fish to Escape Predation under Low Doses of Exposure." Sci Rep **9**(1): 10530.

Sandy, M. and A. Butler (2009). "Microbial Iron Acquisition: Marine and Terrestrial Siderophores." Chemical reviews **109**(10): 4580-4595.

Santini, A., L. Ghelardini, C. De Pace, M. L. Desprez-Loustau, P. Capretti, A. Chandelier, T. Cech, D. Chira, S. Diamandis, T. Gaitniekis, J. Hantula, O. Holdenrieder, L. Jankovsky, T. Jung, D. Jurc, T. Kirisits, A. Kunca, V. Lygis, M. Malecka, B. Marcais, S. Schmitz, J. Schumacher, H. Solheim, A. Solla, I. Szabò, P. Tsopelas, A. Vannini, A. M. Vettraino, J. Webber, S. Woodward and J. Stenlid (2013). "Biogeographical Patterns and Determinants of Invasion by Forest Pathogens in Europe." New Phytologist **197**(1): 238-250.

Santos-Montañez, J., J. A. Benavides-Montaño, A. K. Hinz and V. Vadyvaloo (2015). "Yersinia Pseudotuberculosis Ip32953 Survives and Replicates in Trophozoites and Persists in Cysts of *Acanthamoeba Castellanii.*" FEMS Microbiol Lett **362**(13): fnv091-fnv091.

Sasikaran, J., M. Ziemski, P. K. Zadora, A. Fleig and I. A. Berg (2014). "Bacterial Itaconate Degradation Promotes Pathogenicity." Nat Chem Biol **10**(5): 371-377.

Schaack, S., C. Gilbert and C. Feschotte (2010). "Promiscuous DNA: Horizontal Transfer of Transposable Elements and Why It Matters for Eukaryotic Evolution." Ecol Evol **25**(9): 537-546.

Schmidt, R., V. Cordovez, W. de Boer, J. Raaijmakers and P. Garbeva (2015). "Volatile Affairs in Microbial Interactions." The ISME Journal **9**(11): 2329-2335.

Schmiel, D. H., E. Wagar, L. Karamanou, D. Weeks and V. L. Miller (1998). "Phospholipase a of *Yersinia Enterocolitica* Contributes to Pathogenesis in a Mouse Model." Infection and Immunity **66**(8): 3941-3951.

Schwyn, B. and J. B. Neilands (1987). "Universal Cas Assay for the Detection and Determination of Siderophores." Analytical biochemistry **160**: 47-56.

Seeman, T. (2015). " Snippy: Fast Bacterial Variant Calling from Ngs Reads."

Seemann, T. (2014). "Prokka: Rapid Prokaryotic Genome Annotation." Bioinformatics **30**(14): 2068-2069.

Selin, C., R. Habibian, N. Poritsanos, S. N. P. Athukorala, D. Fernando and T. R. De Kievit (2010). "Phenazines Are Not Essential for Pseudomonas Chlororaphis Pa23 Biocontrol of *Sclerotinia Sclerotiorum,* but Do Play a Role in Biofilm Formation." FEMS Microbiol Ecol **71**(1): 73-83.

Sitter, T. L. (2020). Evolutionary Divergence of the Insect Disease-Encoding *Serratia* Plasmid Padap. New Zealand, Lincoln University. **PhD:** 272.

Snipen, L. and K. H. Liland (2015). "Micropan: An R-Package for Microbial Pan-Genomics." BMC Bioinformatics **16**(1): 79.

Soares-da-Silva, J., V. C. S. Pinheiro, E. Litaiff-Abreu, R. A. Polanczyk and W. P. Tadei (2015). "Isolation of *Bacillus Thuringiensis* from the State of Amazonas, in Brazil, and Screening against *Aedes Aegypti* (Diptera, Culicidae)." Revista Brasileira de Entomologia **59**(1): 1-6.

Son, M. S. a. T., R.K (2014). "Genetic Screens and Biochemical Assays to Characterize *Vibrio Cholerae* O1 Biotypes: Classical and El Tor." Curr Protoc Microbiol **22A**(6A2): 2-17.

Sota, M., H. Yano, J. M Hughes, G. W. Daughdrill, Z. Abdo, L. J. Forney and E. M. Top (2010). "Shifts in the Host Range of a Promiscuous Plasmid through Parallel Evolution of Its Replication Initiation Protein." The ISME Journal **4**(12): 1568-1580.

Stalder, T., L. M. Rogers, C. Renfrow, H. Yano, Z. Smith and E. M. Top (2017). "Emerging Patterns of Plasmid-Host Coevolution That Stabilize Antibiotic Resistance." Sci Rep **7**(1): 4853.

Stehr, F., M. Kretschmar, C. Kröger, B. Hube and W. Schäfer (2003). "Microbial Lipases as Virulence Factors." J Mol Catalysis **22**(5): 347-355.

Suh, Y., S. Jin, T. K. Ball and M. J. Benedik (1996). "Two-Step Secretion of the *Serratia Marcescens* Extracellular Nuclease." J Bacteriol **178**(13): 3771-3778.

Sutton, M. D., B. T. Smith, V. G. Godoy and G. C. Walker (2000). "The Sos Response: Recent Insights into Umudc-Dependent Mutagenesis and DNA Damage Tolerance." Annu Rev Genet **34**(1): 479-497.

Szekeres, S., M. Dauti, C. Wilde, D. Mazel and D. A. Rowe-Magnus (2007). "Chromosomal Toxin–Antitoxin Loci Can Diminish Large-Scale Genome Reductions in the Absence of Selection." Mol Microbiol **63**(6): 1588-1605.

Szewczyk, E., T. Nayak, C. E. Oakley, H. Edgerton, Y. Xiong, N. Taheri-Talesh, S. A. Osmani and B. R. Oakley (2006). "Fusion Pcr and Gene Targeting in Aspergillus Nidulans." Nat Protoc **1**(6): 3111-3120.

Thoma, S. and M. Schobert (2009). "An Improved *Escherichia Coli* Donor Strain for Diparental Mating." FEMS Microbiol Lett **294**(2): 127-132.

Thompson, J. D., T. J. Gibson and D. G. Higgins (2002). "Multiple Sequence Alignment Using Clustalw and Clustalx." Curr Protoc Bioinformatics **Chapter 2**: Unit 2 3.

Thompson, L. J., D. S. Merrell, B. A. Neilan, H. Mitchell, A. Lee and S. Falkow (2003). "Gene Expression Profiling of *Helicobacter Pylori* Reveals a Growth-Phase-Dependent Switch in Virulence Gene Expression." Infec Immun **71**(5): 2643-2655.

Tian, J., N. He, L. Hale, S. Niu, G. Yu, Y. Liu, E. Blagodatskaya, Y. Kuzyakov, Q. Gao and J. Zhou (2018). "Soil Organic Matter Availability and Climate Drive Latitudinal Patterns in Bacterial Diversity from Tropical to Cold Temperate Forests." Funct Ecol **32**(1): 61-70.

Townsend, R. J. (2002). "Grass Grub Lifecycle and Control." Cereals 109. 2020.

Trought, T. E. T., T. A. Jackson and R. A. French (1982). "Incidence and Transmission of a Disease of Grass Grub (*Costelytra Zealandica*) in Canterbury." New Zealand Journal of Experimental Agriculture **10**(1): 79-82.

Úbeda, C., E. Maiques, E. Knecht, Í. Lasa, R. P. Novick and J. R. Penadés (2005). "Antibiotic-Induced Sos Response Promotes Horizontal Dissemination of Pathogenicity Island-Encoded Virulence Factors in *Staphylococci.*" Molecul Microbiol **56**(3): 836-844.

UniProt, C. (2008). "The Universal Protein Resource (Uniprot)." Nucleic Acids Res **36**(Database issue): D190-D195.

Vallet-Gely, I., B. Lemaitre and F. Boccard (2008). "Bacterial Strategies to Overcome Insect Defences." Nat Rev Microbiol **6**(4): 302-313.

van de Guchte, M. (2017). "Horizontal Gene Transfer and Ecosystem Function Dynamics." Trends Microbiol **25**(9): 699-700.

van der Straat, L., M. Vernooij, M. Lammers, W. van den Berg, T. Schonewille, J. Cordewener, I. van der Meer, A. Koops and L. H. de Graaff (2014). "Expression of the *Aspergillus Terreus* Itaconic Acid Biosynthesis Cluster in *Aspergillus Niger*." Microbial Cell Factories **13**(1): 11.

Velappan, N., D. Sblattero, L. Chasteen, P. Pavlik and A. R. M. Bradbury (2007). "Plasmid Incompatibility: More Compatible Than Previously Thought?" Protein Eng Des Sel **20**(7): 309-313.

Vigil-Stenman, T., K. Ininbergs, B. Bergman and M. Ekman (2017). "High Abundance and Expression of Transposases in Bacteria from the Baltic Sea." The ISME Journal **11**(11): 2611-2623.

Vijverberg, H. P., J. M. van der Zalm and J. van der Bercken (1982). "Similar Mode of Action of Pyrethroids and Ddt on Sodium Channel Gating in Myelinated Nerves." Nature **295**(5850): 601-603.

Vollmer, W. and A. Tomasz (2002). "Peptidoglycan N-Acetylglucosamine Deacetylase, a Putative Virulence Factor in *Streptococcus Pneumoniae*." Infect Immun **70**(12): 7176-7178.

Waack, S., O. Keller, R. Asper, T. Brodag, C. Damm, W. F. Fricke, K. Surovcik, P. Meinicke and R. Merkl (2006). "Score-Based Prediction of Genomic Islands in Prokaryotic Genomes Using Hidden Markov Models." BMC Bioinformatics **7**(1): 142.

Walker, B. J., T. Abeel, T. Shea, M. Priest, A. Abouelliel, S. Sakthikumar, C. A. Cuomo, Q. Zeng, J. Wortman, S. K. Young and A. M. Earl (2014). "Pilon: An Integrated Tool for Comprehensive Microbial Variant Detection and Genome Assembly Improvement." PLoS One **9**(11): e112963.

Wang, X., Y. Li, H. Jing, Y. Ren, Z. Zhou, S. Wang, B. Kan, J. Xu and L. Wang (2011). "Complete Genome Sequence of a *Yersinia Enterocolitica* "Old World" (3/O:9) Strain and Comparison with the "New World" (1b/O:8) Strain." J Clin Microbiol **49**(4): 1251-1259.

Williams, L. E. and J. J. Wernegreen (2015). "Genome Evolution in an Ancient Bacteria-Ant Symbiosis: Parallel Gene Loss among *Blochmannia* Spanning the Origin of the Ant Tribe Camponotini." PeerJ **3**: e881.

Woodcock, B. A., L. Ridding, S. N. Freeman, M. G. Pereira, D. Sleep, J. Redhead, D. Aston, N. L. Carreck, R. F. Shore, J. M. Bullock, M. S. Heard and R. F. Pywell (2018). "Neonicotinoid Residues in Uk Honey Despite European Union Moratorium." PLoS One **13**(1): e0189681.

Xue, C. and D. G. Sashital (2019). "Mechanisms of Type I-E and I-F Crispr-Cas Systems in *Enterobacteriaceae*." EcoSal Plus **8**(2): 10.1128/ecosalplus.ESP-0008-2018.

Yang, Z., Y. Zhang, E. K. Wafula, L. A. Honaas, P. E. Ralph, S. Jones, C. R. Clarke, S. Liu, C. Su, H. Zhang, N. S. Altman, S. C. Schuster, M. P. Timko, J. I. Yoder, J. H. Westwood and C. W. dePamphilis (2016). "Horizontal Gene Transfer Is More Frequent with Increased Heterotrophy and Contributes to Parasite Adaptation." Proc Natl Acad Sci U S A **113**(45): E7010-E7019.

Yanisch-Perron, C., J. Vieira and J. Messing (1985). "Improved M13 Phage Cloning Vectors and Host Strains: Nucleotide Sequences of the M13mp18 and Puc19 Vectors." Gene **33**(1): 103-119.

Yin, X. and G. Stotzky (1997). Gene Transfer among Bacteria in Natural Environments. Adv Appl Microbiol. S. L. Neidleman and A. I. Laskin, Academic Press. **45:** 153-212.

Zagaglia, C., M. Casalino, B. Colonna, C. Conti, A. Calconi and M. Nicoletti (1991). "Virulence Plasmids of Enteroinvasive *Escherichia Coli* and *Shigella Flexneri* Integrate into a Specific Site on the Host Chromosome: Integration Greatly Reduces Expression of Plasmid-Carried Virulence Genes." Infect Immun **59**(3): 792-799.

Zdzalik, M., M. Kalinska, M. Wysocka, J. Stec-Niemczyk, P. Cichon, N. Stach, N. Gruba, H. R. Stennicke, A. Jabaiah, M. Markiewicz, S. Kedracka-Krok, B. Wladyka, P. S. Daugherty, A. Lesner, K. Rolka, A. Dubin, J. Potempa and G. Dubin (2013). "Biochemical and Structural Characterization of Spld Protease from *Staphylococcus Aureus.*" PLoS One **8**(10): e76812.

Zhang, H. and T. A. Jackson (2008). "Autochthonous Bacterial Flora Indicated by Pcr-Dgge of 16s Rrna Gene Fragments from the Alimentary Tract of *Costelytra Zealandica* (Coleoptera: Scarabaeidae)." J Appl Microbiol **105**(5): 1277-1285.

Zhang, Y., J. Li, W. Zhang, H. Shi, F. Luo, Y. Hikichi, X. Shi and K. Ohnishi (2018). "A Putative Lysr-Type Transcriptional Regulator Prho Positively Regulates the Type Iii Secretion System and Contributes to the Virulence of *Ralstonia Solanacearum*." Mol Plant Pathol **19**(8): 1808-1819.

Zydenbos, S. M., R. Townsend, P. M. S. Lane, S. Mansfield, M. O'Callaghan, C. van Koten and T. Jackson (2016). "Effect of *Serratia Entomophila* and Diazinon Applied with Seed against Grass Grub Populations on the North Island Volcanic Plateau." New Zealand Plant Protection **69**: 86-93.

Appendix A

Methods

Below follows the recipes for solutions used prior and described in the methods (Chapter 2).

A.1 Growth medium for IAA auxin detection assay

5g glucose

0.025g yeast extract

0.204g L-tryptophan

1L dd.H_20

Solution was then aliquoted into 10mL McCartney bottles and autoclaved at standard settings.

A.2 Kado and Liu megaplasmid visualisation buffers

The following two solution lists contain the recipe used for buffers in Kado and Liu plasmid visualisation.

A.2.1 Lysis solution

0.121g Tris Base

6mL SDS (10%)

0.82mL 2M NaOH

Adjusted to 20mL with autoclaved dd.H_20

A.2.2 E buffer (2×)

0.968g Tris base

3.3mL 3M NaAc

4mL 0.5M EDTA

Adjusted to 100mL with autoclaved dd.H_20 and adjusted to pH 7.9 with glacial acetic acid (20×E)

To obtain working 2× solution, 20×E buffer is then diluted 1:10 with autoclaved dd.H_20

A.3 Plasmid mini prep solutions

A.3.1 Solution 1

5mL 1M Tris-HCL (pH 8)

4mL 0.5M EDTA (pH 8)

182 mL H_2O

Solution is then autoclaved before 9mL of 20% glucose is added

2.5µL of RNase was then added per 50mL of solution 1

A.3.2 Solution 2

1mL 2M NaOH

1mL 10% SDS

8mL autoclaved dd.H_20

A.4 M9 Minimal Salts medium recipe

A.4.1 M9 salts (5×)

32g $Na_2HPO_4.7H_2O$

7.5g KH_2PO_4

1.25g NaCl

2.5g NH_4Cl

Solution adjusted to 500mL with dd.H_20 before 100mL aliquots were autoclaved at standard settings.

A.4.2 M9 minimal medium

20mL M9 salts

2mL 20% glucose (filter sterilised)

200µL $MgSO_4$ (autoclaved)

10µL $CaCl_2$ (autoclaved)

78mL dd.H_2O (autoclaved)

Appendix B

Chapter 4

Below contains the supplementary data as described in Chapter 4, bacterial phenotyping of the *Serratia* spp.

B.1 Comprehensive pathotype and geographic isolate data

Table B 1 shows the comprehensive isolate list of those used in this study including pathotype and sampling location. Those that are further sequenced and used in this study are also indicated.

Table B 1 Geographic location data for all isolates used in this study
Location is to closest known point. Virulence is associated with results from this and previous studies. Sequenced refers to sequencing via either Illumina or PacBio further used in comparative genomics in this study. Isolation given in Year/ month refers to specific isolate isolation.

Species	Location	Isolate	Virulence	Sequencing	Isolation	Reference
Serratia entomophila	Montalto	158	Chronic	Illumina	1988-06	
Serratia entomophila	Ashbury, Canterbury	176	Chronic	Illumina	1990-11	(Dodd 2003)
Serratia entomophila	Carew	192	Non-path		1988-06	(Dodd 2003)
Serratia entomophila	Timaru	204	Chronic		1992-01	(Dodd 2003)
Serratia entomophila	Timaru	207	Chronic		1988-06	(Dodd 2003)
Serratia entomophila	Oamaru	210	non-path	Illumina	1988-06	(Dodd 2003)
Serratia entomophila	Reefton, West Coast	219	Non-path	PacBio	1992-01	
Serratia entomophila	Pasteur institute, France	220	Non-path	Illumina	1992-01	(Dodd 2003)
Serratia entomophila	Chatham Islands	294	Non-path	Illumina	1989-01	
Serratia entomophila	Chatham Islands	295	Non-path		1989-01	
Serratia entomophila	Waingaro, Waikato	305	Chronic	Illumina	1989-04	
Serratia entomophila	Carterton, Wairarapa	306	Non-path		1989-04	
Serratia entomophila	Takapau, Hawkes Bay	307	Chronic		1992-01	(Dodd 2003)
Serratia entomophila	Putaruru, Waikato	311	Non-path		1989-04	
Serratia entomophila	Landcorp, Taupo	314	Non-path		1989-04	(Dodd 2003)

Serratia entomophila	Kimbolton, Manawatū	315	Non-path		1989-04	
Serratia entomophila	Motueka, Nelson	328	Non-path		1989-11	(Dodd 2003)
Serratia entomophila	Motueka, Nelson	329	Non-path		1989-11	(Dodd 2003)
Serratia entomophila	Ashbury, Canterbury	340	Chronic		1993-06	
Serratia entomophila	Ashbury, Canterbury	341	Chronic		1989-11	
Serratia entomophila	Ashbury, Canterbury	345	Non-path	PacBio	1989-11	(Dodd 2003)
Serratia entomophila	Takapau, Hawkes Bay	359	Chronic		1989-12	
Serratia entomophila	Southland	364	Chronic	Illumina	1990-04	
Serratia entomophila	Canterbury, Ashbury	398	Chronic	Illumina	1991-02	
Serratia entomophila	Chatham Islands	399	Non-path		1991-02	
Serratia entomophila	Motueka, Nelson	400	Non-path		1991-02	(Dodd 2003)
Serratia entomophila	Motueka, Nelson	401	Non-path		1991-02	
Serratia entomophila	Chatham Islands	438	Non-path		1991-07	
Serratia entomophila	Chatham Islands	440	Non-path	PacBio	1991-07	(Dodd 2003)
Serratia entomophila	Chatham Islands	442	Non-path	Illumina	1991-07	
Serratia entomophila	Takapau, Hawkes Bay	451	Chronic		1991-07	
Serratia entomophila	Fairton, Canterbury	468	Chronic		1991-07	
Serratia entomophila	Canterbury	477	Non-path	Illumina	1991-07	
Serratia entomophila	Tihoi	480	Chronic		1991-07	(Dodd 2003)
Serratia entomophila	Motueka, Nelson	482	Non-path	Illumina	1991-07	
Serratia entomophila	Fairton, Canterbury	485	Chronic		1991-07	
Serratia entomophila	Canterbury, Ashbury	498	Non-path		1992-01	
Serratia entomophila	Gore, Southland	562	Chronic		1992-05	
Serratia entomophila	Manawatū-Wanganui, Rangitikei	566	Non-path		1992-05	
Serratia entomophila	Canterbury	625	Non-path		1993-06	(Dodd 2003)
Serratia entomophila	Canterbury, Fairton	626	Chronic	PacBio/ Illumina	1993-06	(Dodd 2003)
Serratia entomophila	Waikato	766	Non-path		1994-07	
Serratia entomophila	Canterbury	1021	Non-path			

Serratia entomophila	Canterbury	1086	Non-path		1997-07	
Serratia entomophila	Hawkes Bay	1100	Chronic	Illumina	1997-08	(Dodd 2003)
Serratia entomophila	Waikato, Hinuera	1235	Non-path		1998-04	
Serratia entomophila	Southland	1343	Non-path	Illumina	1998-07	
Serratia entomophila	Waikato, Hinuera	1360	Non-path		1998-07	
Serratia entomophila	Canterbury, Culverden,	1373	Non-path		1998-07	
Serratia entomophila	Hawkes Bay	1554	Non-path		1999-01	
Serratia entomophila	Southland Gore	1651	Chronic		1999-04	
Serratia entomophila	Canterbury	iDIA	Chronic	Illumina		
Serratia entomophila	Waikato, Tokoroa	I	Non-path	Illumina		
Serratia entomophila	Manawatū-Whanganui, Rangitikei	Man3	Chronic	Illumina		
Serratia entomophila	Cape Fowlwind, West Coast	MC2	Non-path	Illumina		
Serratia entomophila	Canterbury, Fairton	A1M02	Chronic	Illumina		
Serratia entomophila	West Coast, Cape Fowlwind	Sent1	Non path	Illumina		
Serratia proteamaculans	Southland, Te Anau	4	Chronic	Illumina		(Sitter 2020)
Serratia proteamaculans	Manawatū-Whanganui, Tihoi, Taupo	142	Chronic	PacBio/ Illumina		(Dodd 2003)
Serratia proteamaculans	Methven, Canterbury	143	Chronic	PacBio/ Illumina		(Dodd 2003)
Serratia proteamaculans	Waimate, South Canterbury	145	Chronic	Illumina		(Dodd 2003)
Serratia proteamaculans	Methven, Canterbury	149	Non-path	PacBio/ Illumina		
Serratia proteamaculans	Tasman, Motueka	299	Chronic	Illumina		(Dodd 2003)
Serratia proteamaculans	Tasman, Motueka	336	Chronic	Illumina		(Dodd 2003)
Serratia liquefaciens	Manawatū-Whanganui, Taupo	376	Chronic	Illumina		(Dodd 2003)
Serratia liquefaciens	Manawatū-Whanganui, Taupo	377	Non-path	Illumina		
Serratia ficaria	France	457	Chronic	Illumina		
Serratia proteamaculans	Hawkes bay	465	Non-path	Illumina		
Serratia proteamaculans	Argentina	495	Non-path			(Dodd 2003)
Serratia proteamaculans	Canterbury	1048	Chronic	PacBio		(Dodd 2003)

Serratia proteamaculans	Canterbury	1071	Chronic	Illumina		(Dodd 2003)
Serratia proteamaculans	Hawkes bay	1129	Chronic	Illumina		(Dodd 2003)
Serratia proteamaculans	Canterbury	1137	Non-path	Illumina		
Serratia proteamaculans	Auckland	1475	Chronic	Illumina		(Dodd 2003)
Serratia marcescens	Canterbury, St Albans	1577	Non-path	Illumina		
Serratia proteamaculans	Manawatū-Whanganui, Taupo	1769	Chronic	Illumina		
Serratia proteamaculans	Manawatū-Whanganui, Rua pehu	3041	non-path	Illumina		
Serratia proteamaculans	Southland, Te Anau	12a	Chronic	Illumina		
Serratia proteamaculans	Canterbury, Methven	12newD	Chronic	Illumina		
Serratia proteamaculans	Canterbury, Lake Coleridge	20E	Chronic			
Serratia proteamaculans	Canterbury, Lake Coleridge	25E	Chronic	Illumina		
Serratia proteamaculans	Canterbury, Methven	28F	Chronic			
Serratia proteamaculans	Canterbury, Lincoln	AGR96X	Hyper	PacBio/Illumina		(Hurst et al. 2018)
Serratia proteamaculans		CfB	Chronic	Illumina		
Serratia proteamaculans	Canterbury, Hororata	F	Chronic			
Serratia proteamaculans	Canterbury, Lake Coleridge	LC	Hyper	Illumina		
Serratia proteamaculans	Canterbury, Lincoln	20093	Hyper	Illumina		
Serratia proteamaculans	Canterbury, Lincoln	10Novel	Hyper	Illumina		
Serratia proteamaculans	Canterbury, Lake Coleridge	MH5	Hyper	Illumina		
Serratia liquefaciens	Canterbury, Waimakariri	F28	Chronic	Illumina		
Serratia proteamaculans	Rakaia, Canterbury	G	Chronic	Illumina		
Serratia proteamaculans	Manawatū-Wanganui, Ruapehu	K	Chronic	Illumina		
Serratia proteamaculans	Manawatū-Wanganui, Ruapehu	L	Chronic	Illumina		
Serratia proteamaculans	Manawatū-Wanganui, Ruapehu	M	Chronic	Illumina		
Serratia proteamaculans		Puna18	Non-path	Illumina		
Yersinia frederiksenii		Y49	Non-path	Illumina		

B.2 Bioassay data

Full bioassay results data is attached in an excel file. Named Appendix B2- Bioassay result data.xlsx

B.3 Full plate assay results

Full plate assay results data is attached in an excel file. Named Appendix B3- Plate result data.xlsx. Each tabbed excel page refers to a plate assay type assessed in Chapters 4 and 7.

Appendix C

Chapter 5

Below is the supplementary data and attached file information as described in Chapter 5, defining variation in the *Serratia entomophila* genome.

C.1 ANI matrix including *Serratia ficaria* isolate 457

An excel file containing the ANI matrix for all *S. entomophila* isolates with addition *S. ficaria* is attached as Appendix C1- ANI matrix with *Serratia ficaria* 457.xlsx

C.2 Roary output file

The Roary output data as a comma delineated .csv file is found attached under the name Appendix C2- Roary output file.csv

C.3 Scoary output files

Scoary output files for both pathotype and Chatham Islands are attached as two separate files

- Appendix C3- Scoary output Chatham Islands.csv
- Appendix C3- Scoary output pathotype.csv

C.4 Comparative genome statistics including *Serratia ficaria* 457

Below is the comparative genome statistics created using the Roary output file and Roary process R studio packaged script (Figure C1).

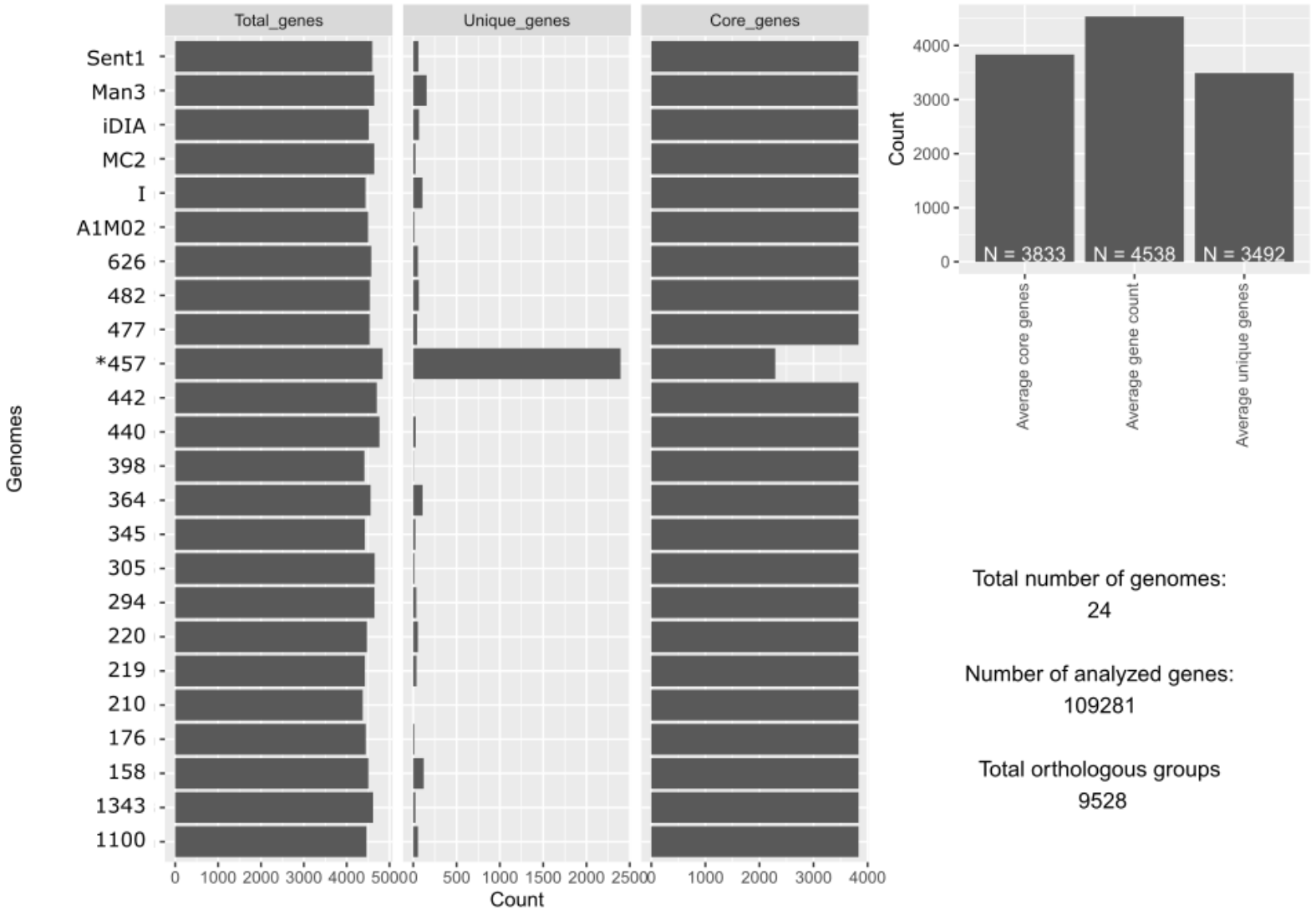


Figure C 1 Comparative genome statistics of the core genome of *Serratia entomophila* with *Serratia ficaria* isolate 457, defined by Roary

C.5 Core genome alignment of only *Serratia entomophila*

Below is a presence/absence heat map of the core and accessory genome of *Serratia entomophila.* This complements the one displayed in Chapter 6, showing only *S. entomophila* without *S. ficaria* (Figure C4).

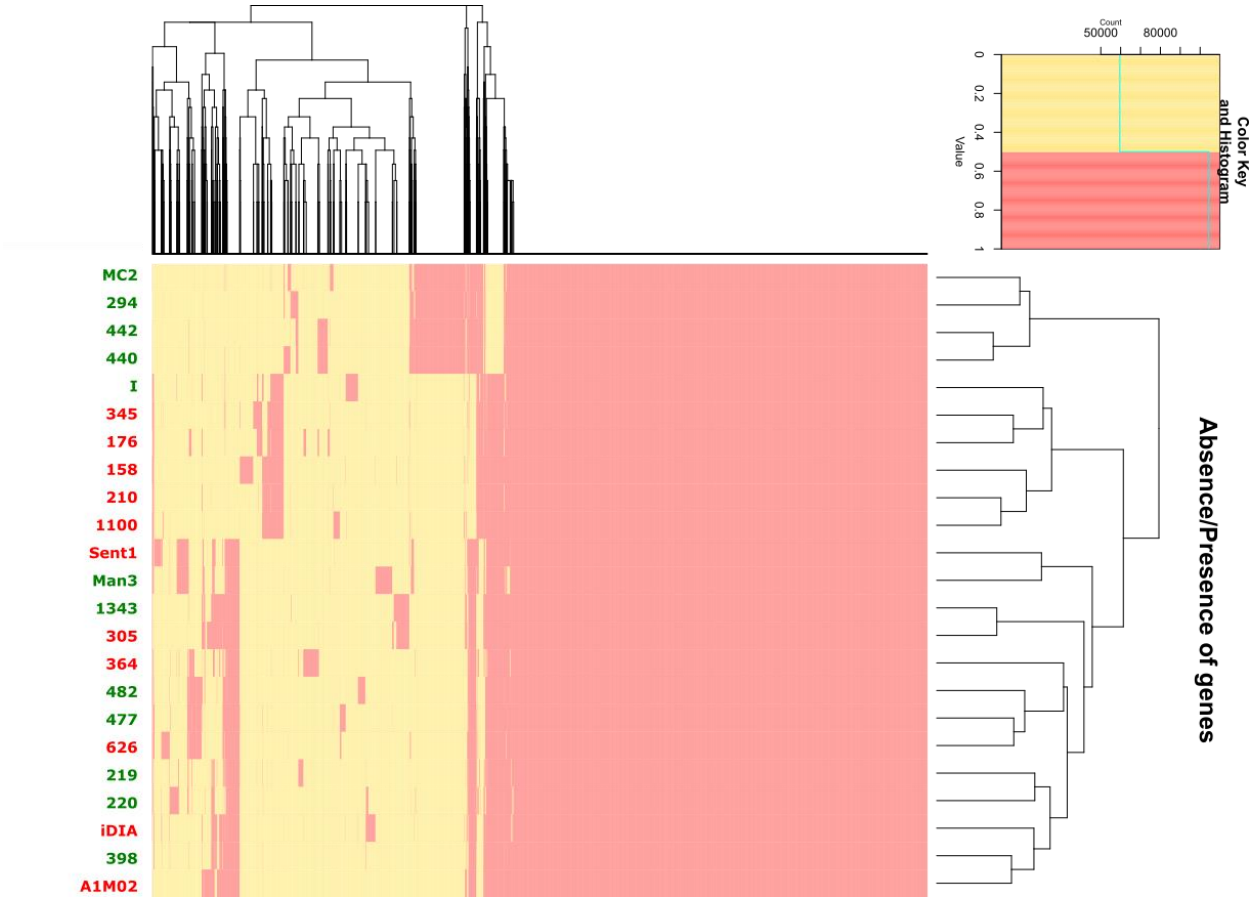


Figure C 2 Core genome alignment of all chromosomes of *Serratia entomophila*
Pathogenic isolates highlighted in red isolate number, non-pathogenic in green.

C.6 Annotation of 40 Kb region isolate 626

Table C 1 shows the annotation using uniprot BLASTp function of the 40 Kb region found in isolate 626 in Chapter 5, section 5.2.7. Percentage identity is displayed as the closest match as per results on Uniprot.

Table C 1 BlastP using Uniprot of the 40Kb chromosomal region unique to *Serratia entomophila* isolate 626

Annotation	BlastP	Percentage Identity	Accession number
Hypothetical	Ogr/Delta-like zinc finger	69.0% *Rahnella aquatilis*	H2IRJ9
Hypothetical	Late control protein D	87.0% *Asticcacaulis benevestitus*	V4NTB6
Hypothetical	Tail assembly protein	89.4% *Asticcacaulis benevestitus*	V4NN27
Hypothetical	Tail protein	64.3% *Asticcacaulis benevestitus*	V4N3Q7
Hypothetical	Putative phage tail protein	70.5% *Ewingella americana* ATCC 33852	A0A085GNP9
Hypothetical	Major tail tube protein	98.2% *Asticcacaulis benevestitus*	V4N072
gpFI CDS	Phage tail sheath monomer	82.5% *Ewingella americana* ATCC 33852	A0A085GNQ1
Hypothetical	Phage tail protein	60.3% *Sodalis* sp. TME1	A0A1J7CAI0
Hypothetical	Phage tail fiber repeat protein	56.5% *Serratia odorifera* DSM 4582	D4DZ45
Hypothetical	Tail protein	89.1% *Asticcacaulis benevestitus*	V4Q9M6
Hypothetical	Baseplate assembly protein	87.2% *Asticcacaulis benevestitus*	V4NLE4
Hypothetical	GPW_gp25 domain-containing protein	74.1% *Sodalis* sp. 159R	A0A4R1N8P2
Hypothetical	Baseplate assembly protein	80.7% *Asticcacaulis benevestitus*	V4Q7P5
Hypothetical	Uncharacterized protein	83.9% *Gibbsiella quercinecans*	A0A250B332
Hypothetical	Phage virion morphogenesis protein	60.3% *Candidatus Erwinia dacicola*	A0A328TQI5
Hypothetical	Phage tail completion protein	66.7% *Erwinia* phage ENT90	F1BUP9
Hypothetical	LysB family protein	42.0% *Buttiauxella agrestis*	A0A085GA11
rrrD 3 CDS	Lysozyme	70.7% *Chimaeribacter arupi*	A0A2N5EJ41
Hypothetical	Uncharacterized protein	69.95% *Serratia ficaria*	A0A240B104
Hypothetical	Tail protein X	92.5% *Asticcacaulis benevestitus*	V4NFD4
Hypothetical	Head completion/stabilization protein	87% *Asticcacaulis benevestitus*	V4MZF5
Hypothetical	Small terminase subunit	76.1% *Sodalis* sp. 159R	A0A4R1N8Z6
Hypothetical	Major capsid protein	78.1% *Erwinia* phage ENT90	F1BUQ8
Hypothetical	Phage capsid scaffolding protein (GPO) serine peptidase	53% *Rahnella aquatilis*	H2IRH5
Hypothetical	Oxidoreductase	83.7% *Hafnia alvei* FB1	A0A097QXC1
Hypothetical	Presumed portal vertex protein	80.2% *Erwinia gerundensis*	A0A0U5L1L9
Hypothetical	Uncharacterized protein	70.4% *Pragia fontium*	A0A0G3CPH6

Hypothetical	Uncharacterized protein	84.0% *Pragia fontium*	A0A0G3CRP5
aplIR CDS	Putative restriction endonuclease	70.7% *Burkholderia glumae*	C5AP23
paeR7IM CDS	Site-specific DNA-methyltransferase (adenine-specific)	63.7% *Idiomarina tyrosinivorans*	A0A432ZT90
Hypothetical	Phage replication protein	59.3% *Buttiauxella agrestis*	A0A085GA27
Hypothetical	RNA-binding protein	57.7% Enterobacter sp. R4-368	R9VFQ2
Hypothetical	Putative MFS family arabinose efflux permease	62.5% *Sulfurirhabdus autotrophica*	A0A4R3XXM2
ybiI 2 CDS	DksA C4-type domain-containing protein	78.6% *Asticcacaulis benevestitus*	V4NNG2
Hypothetical	Uncharacterized protein	86.2% *Asticcacaulis benevestitus*	V4N6L1
Hypothetical	Uncharacterized protein	59.8% *Erwinia* sp. Leaf53	A0A0Q4MNZ1
Hypothetical	Regulator	98.5% *Asticcacaulis benevestitus*	V4Q572
Hypothetical	Uncharacterized protein	60.3% *Serratia nematodiphila* DZ050	A0A086G9S0
xerD 4 CDS	Phage integrase family protein	84.0% *Candidatus Erwinia dacicola*	A0A328TIV5

C.7 Annotation of 33 Kb region isolate 477

Table C 2 shows the annotation using uniprot BLASTp function of the 33 Kb region found in isolate 477 in Chapter 5, section 5.2.7. Percentage identity is displayed as the closest match as per results on Uniprot.

Table C 2 BlastP using Uniprot of the 33Kb chromosomal region unique to *Serratia entomophila* isolate 477

Annotation	BlastP	Percentage Identity	Accession number
xerD_1	Phage integrase	56.8% *Candidatus Sodalis pieranton*	W0HP41
Hypothetical	Uncharacterized protein	55.1% *Candidatus Fukatsuia symbiotica*	A0A2U8IA78
Hypothetical	Uncharacterized protein	42.5% *Serratia* sp. S1B	A0A2V1HMC6
Hypothetical	TerB domain-containing protein	91.4% *Serratia odorifera* DSM 4582	Q1MVI3
Hypothetical	Helix-turn-helix protein	25.8% *Paenibacillus methanolicus*	A0A5S5CLG6
Hypothetical	Helix-turn-helix protein	27.6% *Paenibacillus methanolicus*	A0A653FUC2
Hypothetical	DUF3850 domain-containing protein	55.9% *Serratia* sp. S1B	A0A2V1HG07
Hypothetical	Nucleoid occlusion protein	38.6% *Paludisphaera borealis*	A0A1U7CI59
dam_2	Site-specific DNA-methyltransferase (adenine-specific)	54.1% *Yersinia ruckeri*	W0L905
cysH_1	Phosphoadenosine phosphosulfate reductase family protein	70.8% *Serratia odorifera* DSM 4582	A4W8K0
Hypothetical	Replication endonuclease	56.4% *Serratia sp*. S1B	A0A2V1HJ88
Hypothetical	Phage DNA methylase	69.2% *Pectobacterium atrosepticum*	Q6D3W8
Hypothetical	Uncharacterized protein	27.1% *Bordetella* genomosp. 4	A0A261U5P3
Hypothetical	SGNH/hydrolase family protein	28.4% *Klebsiella* phage ST147-VIM1p	A0A482J319
Hypothetical	Portal protein	64.85% *Salmonella* phage SW5	A0A653FU82
Hypothetical	Putative terminase, ATPase subunit	59.5% *Xenorhabdus doucetiae*	A0A068QV69
Hypothetical	Probable capsid scaffolding protein	44.3% *Dickeya dadantii*	E0SAM7
Hypothetical	Phage major capsid protein	61.4% *Escherichia* phage ESSI2_ev040	A0A653FVA0
Hypothetical	Probable terminase, endonuclease subunit	58.9% *Dickeya dadantii* (strain 3937)	A0A2V1HIC7
Hypothetical	Head completion protein	51.0% *Candidatus Erwinia dacicola*	A0A1E7YZV0
Hypothetical	Head completion protein	51.5% *Candidatus Erwinia dacicola*	A0A1E7YZV0
Hypothetical	Uncharacterized protein	57.85 *Salmonella* phage SW5	A0A5C0CGU0
Hypothetical	DUF2586 family protein	64.3% *Salmonella* phage SW5	A0A5C0CGW1
Hypothetical	Protein of uncharacterized function (DUF2597)	60.9% *Citrobacter koseri*	A0A3D9U994
Hypothetical	Holin	67.4% *Escherichia* phage ESSI2_ev040	A0A653FU89
Hypothetical	Holin	67.4% *Escherichia* phage ESSI2_ev040	A0A653FU89
Hypothetical	Uncharacterized protein	51.3% *Salmonella* phage SW5	A0A5C0CHU1
Hypothetical	Uncharacterized protein	65.1% *Izhakiella* sp. KSNA2	W0HZD8
Hypothetical	Phage tail tape measure protein, TP901 family, core region	46.1% *Enterovibrio nigricans*	A0A2C8FDB6

Hypothetical	Uncharacterized protein	60.4% *Dickeya dadantii*	A0A5C0CGX1
Hypothetical	Putative bacteriophage protein	62.6% *Dickeya dadantii* (strain 3937)	E0SAP1
Hypothetical	Uncharacterized protein	58.0% *Serratia* sp. S1B	A0A653FTA2
Hypothetical	Probable tail fiber protein	39.9% *Dickeya dadantii* (strain 3937)	E0SAP3
Hypothetical	Uncharacterized protein	36.8% *Izhakiella* sp. KSNA2	A0A4P8YFF3
Hypothetical	Uncharacterized protein	52.2% *Serratia* sp. S1B	A0A2V1HEN5
Hypothetical	Uncharacterized protein	54.5% *Erwinia gerundensis*	A0A0U5GIQ5
cspC_3	Cold shock protein	100% *Serratia symbiotica*	R4I0D8

C.8 Genomic island predictions of *Serratia entomophila*

Genomic island predictions for further isolates of *Serratia entomophila* not included in the main body of text of Chapter 5, section 5.2.9 (Figure C3 and Figure C4).

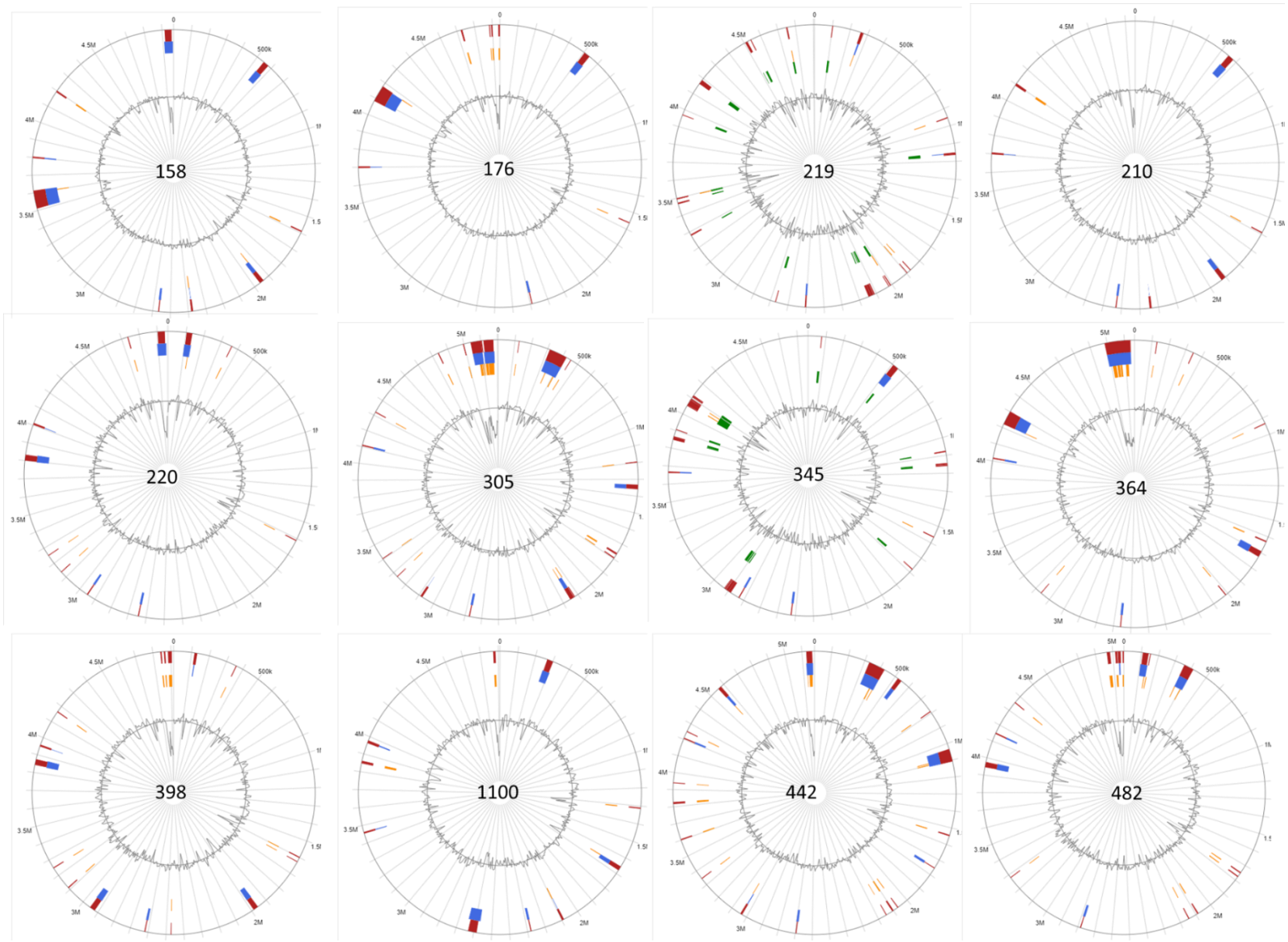


Figure C 3 Predicted genomic islands in isolates of *Serratia entomophila*
Red indicates that the island was predicted by at least one method, where blue (Islandpath-DMOB) green (IslandPick) and orange (SIGI-HMM) indicated individual prediction methods. GC skew shown in the centre circle.

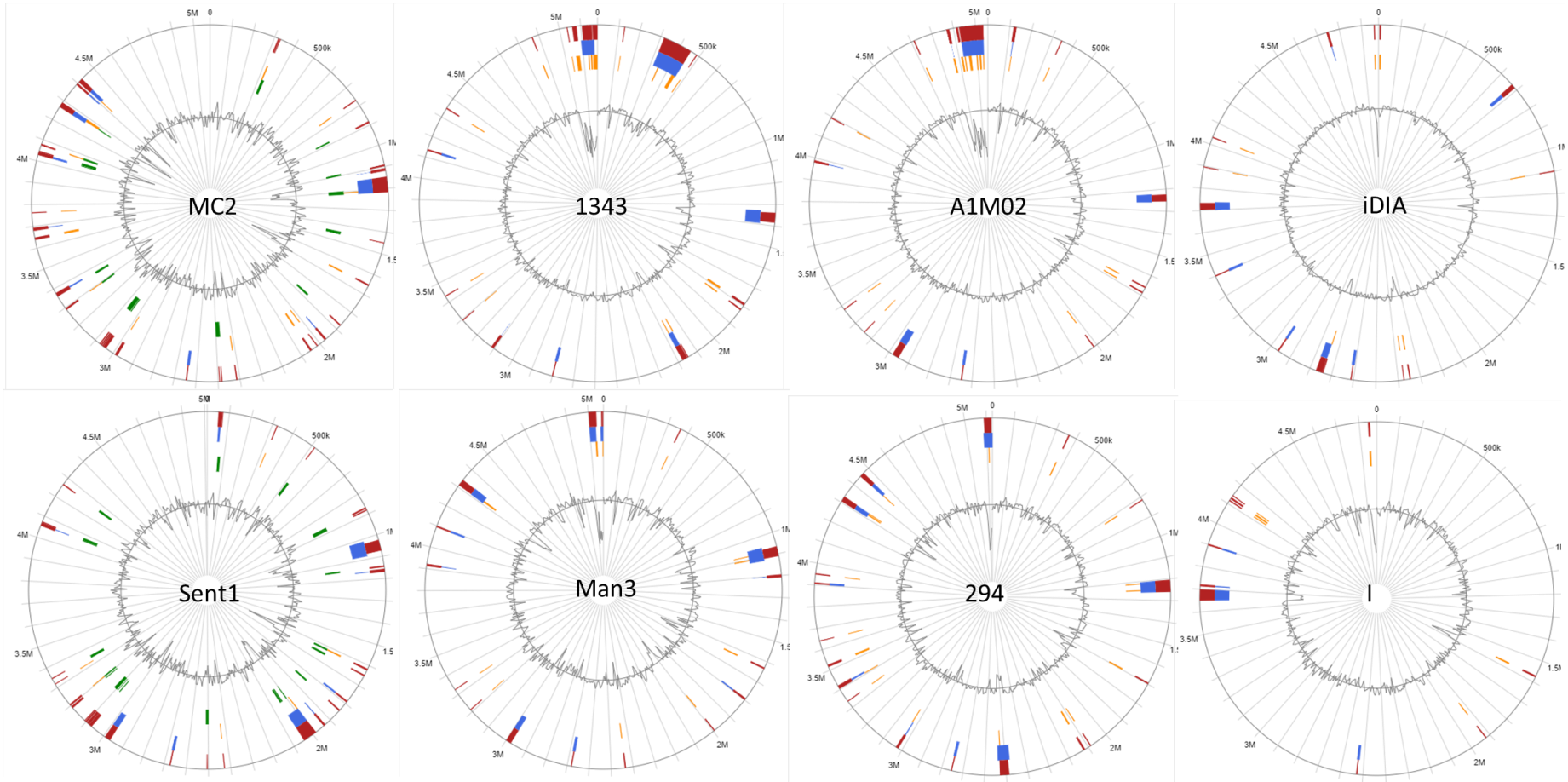


Figure C 4 Predicted genomic islands in isolates of *Serratia entomophila*
Red indicates that the island was predicted by at least one method, where blue (Islandpath-DMOB) green (IslandPick) and orange (SIGI-HMM) indicated individual prediction methods. G+C skew shown in the centre circle.

C.9 Accessory virulence factor motif prediction

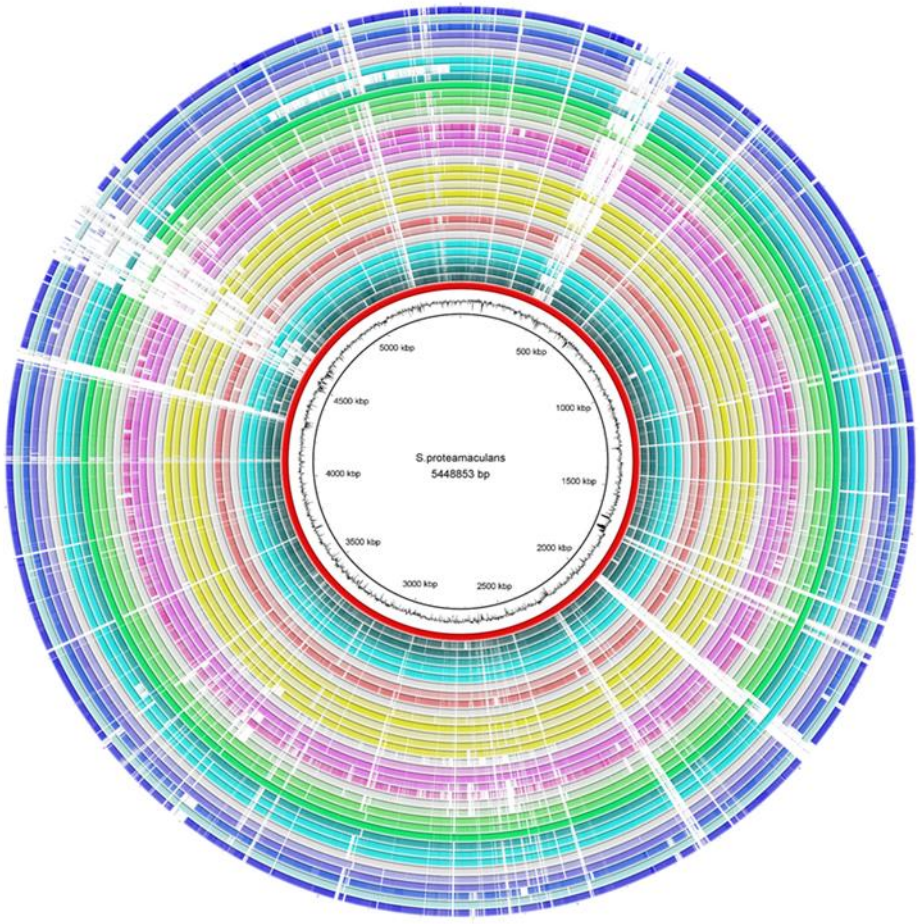
Accessory virulence factor genome predictions were undertaken using hmmsearch as described in Chapter 2, section 2.5.13.2. The merged output files for predictions using hmmsearch for *S. entomophila* isolates are attached as excel files for Pfam family lipases, family 3 lipases and glycoside hydrolases. These files are named as below:

- Appendix C9a- hmmsearch family 3 lipase output entomophila.xlsx
- Appendix C9b- hmmsearch family lipase output entomophila.xlsx
- Appendix C9c- hmmsearch family glycoside hydrolase entomophila.xlsx

Appendix D

Chapter 6

D.1 BRIGS diagrams including isolate keys



Inside to out:
Reference genome 568
Agr96X*
20093*
Novel10*
1129*
LC*
1048
12newD
142
143
145
25e
4
465
CfB
M
MH5
RM5[1]
Sprot
F28[2]
1
1071
12a
1137
1457
149
1769
336
G
K
Puna18
299
3041

Figure D 1 BRIGs diagram with isolate key of *Serratia proteamaculans*
* indicates hypervirulent isolates (inside rings)
[1] indicates RM5, which was not used further in this study.
[2] indicates a later reclassified isolate (reclassified to *Serratia liquefaciens*)

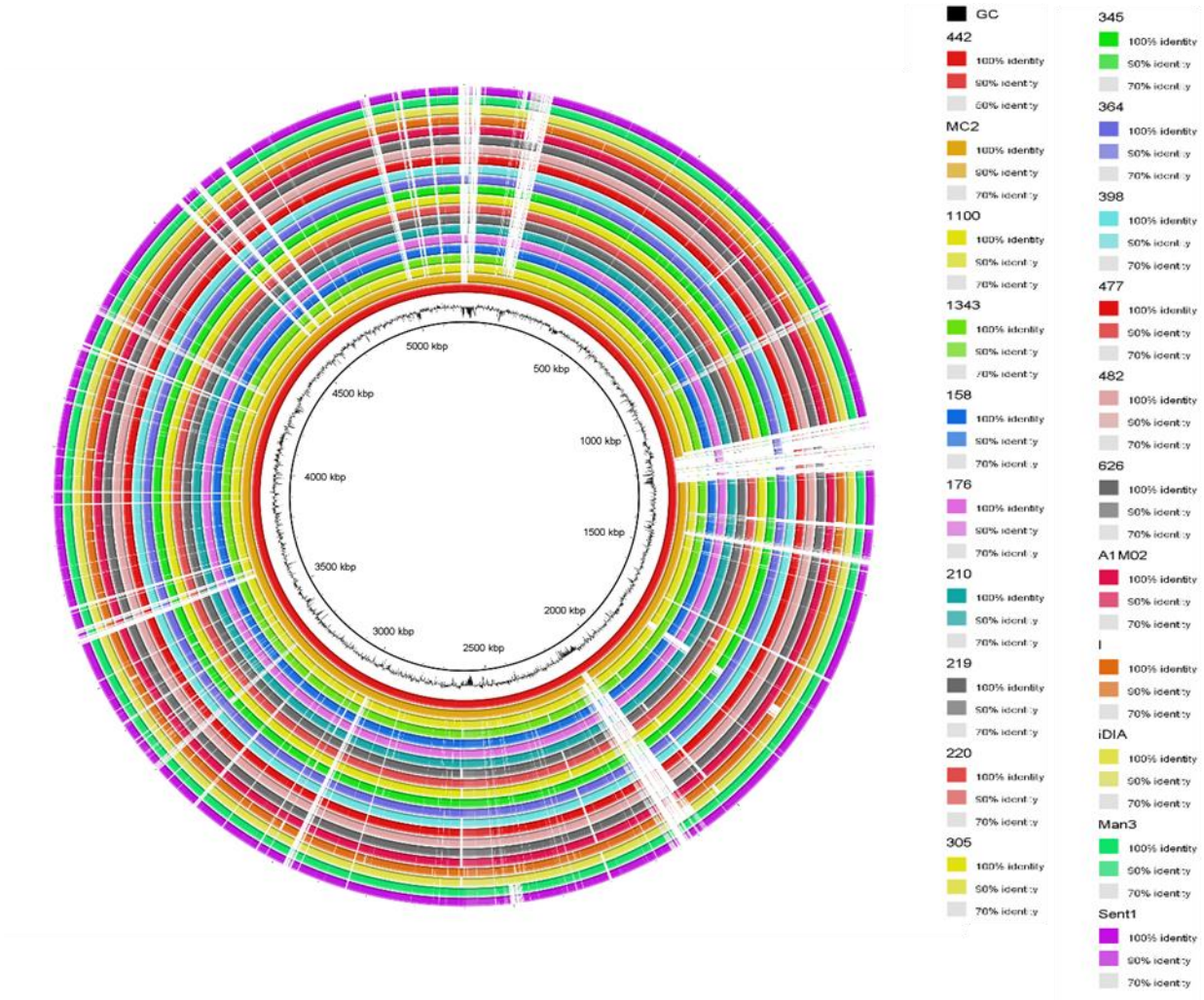


Figure D 2 BRIGs diagram with isolate key of *Serratia entomophila* using Chatham Islands isolate 440 as a reference

D.2 Roary output file

The Roary output data as a comma delineated .csv file is found attached under the name Appendix D2- Roary presence absence all.csv

D.3 Scoary output files

Scoary output files for AGR96X (unique), MEROPs presence absence, pathotype and species separate files

- Appendix D3- Scoary output AGR96X.csv
- Appendix C3- Scoary output MEROPS.csv
- Appendix C3- Scoary output pathotype.csv
- Appendix C3- Scoary output species.csv

D.4 Full KofamKOALA results tables

Table D 1 displays the module results for all isolates not included in the table provided in Chapter 6, section 6.2.4.

Table D 1 Number of complete modules identified by KofamKOALA for metabolic units located on chromosomes of additional *Serratia entomophila*

Pathway module		*S. entomophila*													
		158	**176**	**210**	**219**	**220**	**294**	**305**	**345**	**364**	**398**	**Sent1**	**442**	**482**	**1100**
Carbohydrate metabolism	Central carbohydrate	11	11	11	11	11	11	11	11	11	11	11	11	11	11
	Other carbohydrate	7	6	7	6	6	6	6	6	6	6	6	6	6	6
Energy metabolism	Carbon fixation	1	1	1	1	1	1	1	1	1	1	1	1	1	1
	Nitrogen	1	1	1	1	1	1	1	1	1	1	1	1	1	1
	Sulphur	1	1	1	1	1	1	1	1	1	1	1	1	1	1
	ATP synthesis	5	5	5	5	5	5	5	4	5	5	5	5	5	5
Lipid Metabolism	Fatty acid	4	4	4	4	4	4	4	4	4	4		4	4	4
	Lipid	1	1	1	1	1	1	1	1	1	1	4	1	1	1
Nucleotide metabolism	Purine	3	3	3	3	3	3	3	3	3	3	3	3	3	3
	Pyramidine	2	2	2	2	2	2	2	2	2	2	2	2	2	2
Amino acid metabolism	Serine/ threonine	3	3	3	3	3	3	3	3	3	3	3	3	3	3
	Cysteine/ methionine	4	4	4	5	4	4	5	4	5	5	4	4	4	5
	Branched chain AA	3	3	3	3	3	3	3	3	3	3	3	3	3	3
	Lysine	1	1	1	1	1	1	1	1	1	1	1	1	1	1
	Arginine/ proline	4	4	4	4	4	4	4	5	4	4	4	4	4	4
	Polyamine	1	1	1	1	1	1	1	1	1	1	1	1	1	1
	Histidine	2	2	2	2	2	2	2	2	2	2	2	2	2	2
	Aromatic amino acid	4	4	4	4	4	4	4	4	4	4	4	4	4	4
	Other amino acid	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Glycan metabolism	Lipopolysaccharide	3	3	3	3	3	3	3	3	3	3	3	3	3	3
Metabolism of cofactors and vitamins	Cofactor and vitamin	14	14	14	14	14	15	14	14	15	14	14	15	14	14
Biosynthesis of terpenoids and polyketides	Terpenoid backbone biosynthesis	2	2	2	2	2	2	2	2	2	2	2	2	2	2
	Polyketide sugar unit biosynthesis	1	0	1	1	0	0	1	1	0	1	1	1	1	1
Biosynthesis of other secondary metabolites	Biosynthesis of other bacterial compounds	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Xenobiotics biodegradation	Aromatics degradation	3	3	3	3	3	3	3	3	3	3	3	3	2	3

Table D 2 Number of complete modules identified by KofamKOALA for metabolic units located on chromosomes of selected *Serratia proteamaculans* and *Serratia entomophila*

Pathway module		*S. entomophila*			*S. proteamaculans*								
		1343	**I**	**Man3**	**4**	**142**	**145**	**149**	**299**	**336**	**465**	**1048**	**1129**	**1137**
Carbohydrate metabolism	Central carbohydrate	11	11	11	11	11	11	7	11	9	11	11	9	9
	Other carbohydrate	6	6	6	7	7	7	5	7	7	7	6	7	7
Energy metabolism	Carbon fixation	1	1	1	1	1	1	1	1	1	1	1	1	1
	Nitrogen	1	1	1	1	1	1	0	1	1	1	1	1	1
	Sulphur	1	1	1	1	1	1	1	1	1	1	1	1	1
	ATP synthesis	5	5	5	5	5	5	5	5	5	5	5	5	5
Lipid Metabolism	Fatty acid	4	4	4	4	3	4	5	4	4	4	4	4	4
	Lipid	1	1	1	1	1	1	5	1	1	1	1	1	1
Nucleotide metabolism	Purine	3	3	3	3	3	3	2	3	3	3	3	3	3
	Pyramidine	2	2	2	2	2	2	1	2	2	2	2	2	2
Amino acid metabolism	Serine/ threonine	3	3	2	3	3	3	3	3	3	3	3	3	3
	Cysteine/ methionine	5	4	5	5	4	4	4	4	4	5	4	5	5
	Branched chain AA	3	3	3	3	2	3	1	3	3	3	3	3	3
	Lysine	1	1	1	1	1	1	1	1	1	1	1	1	1
	Arginine/ proline	4	4	4	4	4	4	2	4	4	4	4	4	4
	Polyamine	1	1	1	1	1	1	1	1	1	1	1	2	2
	Histidine	2	2	2	2	1	2	2	2	2	2	2	2	2
	Aromatic amino acid	4	4	4	5	5	5	4	5	5	5	5	5	5
	Other amino acid	1	1	1	1	1	1	1	1	1	1	1	1	1
Glycan metabolism	Lipopolysaccharide	3	3	2	3	3	3	1	3	3	3	3	3	3
Metabolism of cofactors and vitamins	Cofactor and vitamin	14	14	13	16	14	16	13	16	16	16	12	16	16
Biosynthesis of terpenoids and polyketides	Terpenoid backbone biosynthesis	2	2	2	2	2	2	2	2	2	2	2	2	2
	Polyketide sugar unit biosynthesis	1	1	1	2	1	0	0	2	0	0	0	1	0
Biosynthesis of other secondary metabolites	Biosynthesis of other bacterial compounds	1	1	0	0	1	2	2	0	2	2	2	2	1
Xenobiotics biodegradation	Aromatics degradation	3	2	3	1	1	1	1	1	1	1	1	1	1

Table D 3 Number of complete modules identified by KofamKOALA for metabolic units located on chromosomes of selected *Serratia proteamaculans*

Pathway module		*S. proteamaculans*										
		1769	**20093**	**12a**	**12newD**	**25e**	**G**	**K**	**LC**	**M**	**MH5**	**Puna18**
Carbohydrate metabolism	Central carbohydrate	11	11	11	11	11	11	11	9	11	9	11
	Other carbohydrate	7	7	7	7	7	7	7	7	7	7	7
Energy metabolism	Carbon fixation	1	1	1	1	1	1	1	1	1	1	1
	Nitrogen	1	1	1	1	1	1	1	1	1	1	1
	Sulphur	1	1	1	1	1	1	1	1	1	1	1
	ATP synthesis	5	5	5	5	5	5	5	5	5	5	5
Lipid Metabolism	Fatty acid	4	4	4	4	4	4	4	4	4	4	4
	Lipid	1	1	1	1	1	1	1	1	1	1	1
Nucleotide metabolism	Purine	3	3	3	3	3	3	3	3	3	3	3
	Pyramidine	2	2	2	2	2	2	2	2	2	2	2
Amino acid metabolism	Serine/ threonine	3	3	3	3	3	3	3	3	3	3	3
	Cysteine/ methionine	5	5	5	4	4	4	4	5	4	4	5
	Branched chain AA	3	3	3	3	3	3	3	3	3	3	3
	Lysine	1	1	1	1	1	1	1	1	1	1	1
	Arginine/ proline	4	4	4	4	4	4	4	4	4	4	4
	Polyamine	1	1	1	1	1	2	2	1	1	1	1
	Histidine	2	2	2	2	2	2	2	2	2	2	2
	Aromatic amino acid	5	5	5	5	5	5	5	5	5	5	5
	Other amino acid	1	1	1	1	1	1	1	1	1	1	1
Glycan metabolism	Lipopolysaccharide	3	3	3	3	3	3	3	3	3	3	3
Metabolism of cofactors and vitamins	Cofactor and vitamin	16	16	16	16	16	16	16	16	16	15	16
Biosynthesis of terpenoids and polyketides	Terpenoid backbone biosynthesis	2	2	2	2	2	2	2	2	2	2	2
	Polyketide sugar unit biosynthesis	0	1	1	1	0	1	1	0	0	0	1
Biosynthesis of other secondary metabolites	Biosynthesis of other bacterial compounds	2	2	2	2	2	2	2	2	2	2	2
Xenobiotics biodegradation	Aromatics degradation	1	1	1	1	1	1	1	1	1	0	1

D.5 Genomic island predications of *Serratia proteamaculans*

Genomic island predictions for further isolates of *Serratia proteamaculans* not included in the main body of text of Chapter 6, section 6.2.6.1. (Figure D3 and Figure D4).

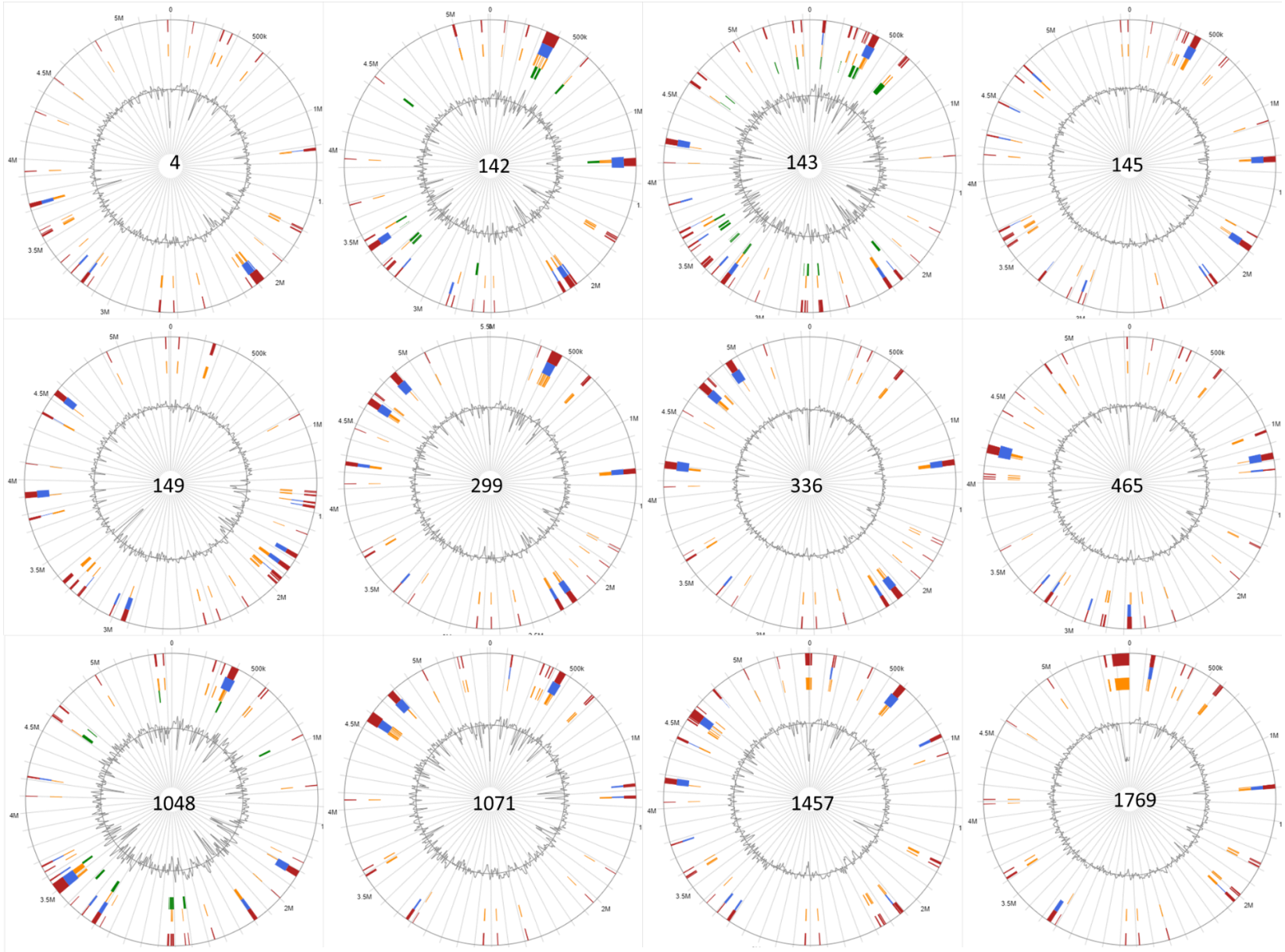


Figure D 3 Predicted genomic islands in isolates of *Serratia proteamaculans*
Red indicates that the island was predicted by at least one method, where blue (Islandpath-DMOB) green (IslandPick) and orange (SIGI-HMM) indicated individual prediction methods. G+C skew shown in the centre circle.

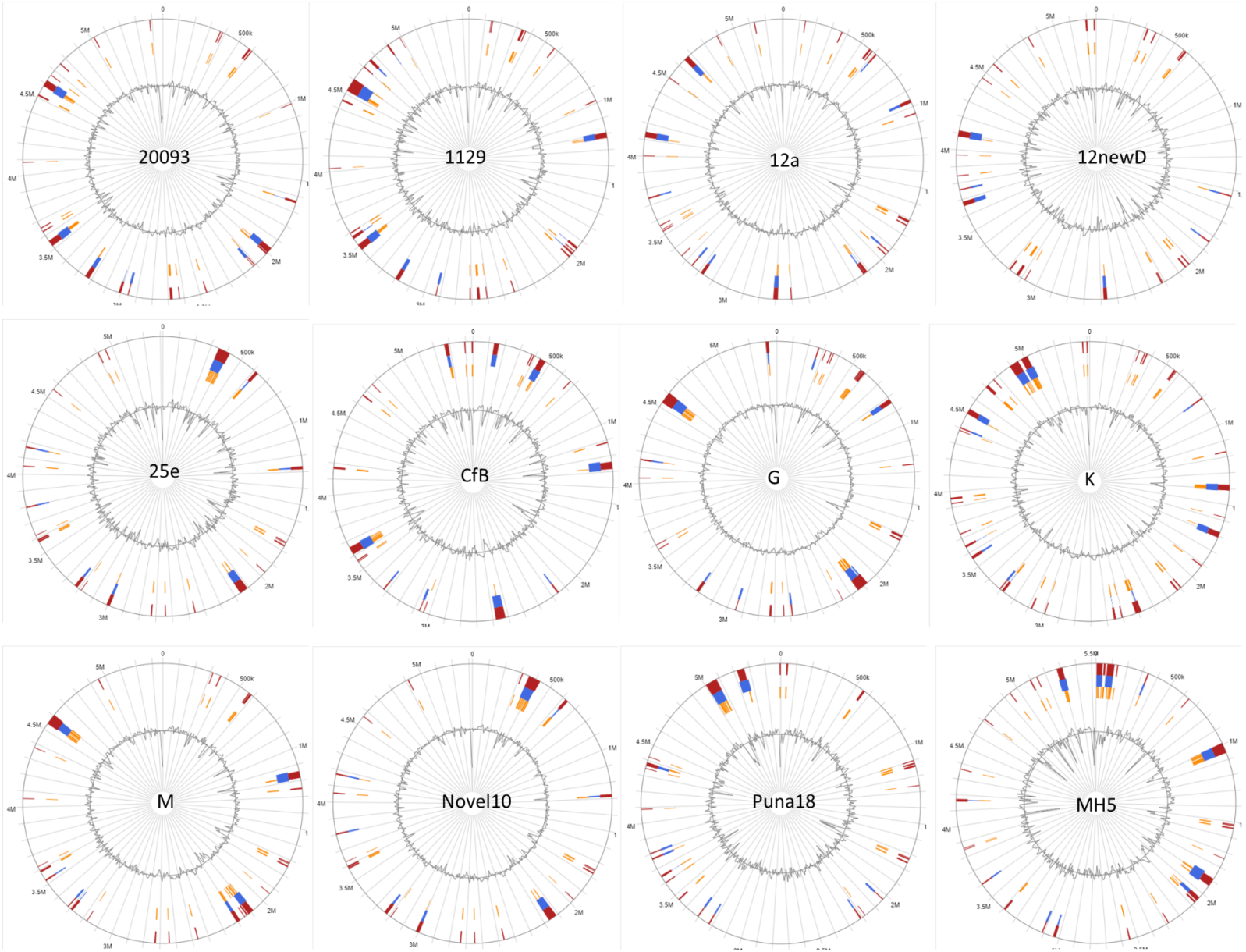


Figure D 4 Predicted genomic islands in isolates of *Serratia proteamaculans*

Red indicates that the island was predicted by at least one method, where blue (Islandpath-DMOB) green (IslandPick) and orange (SIGI-HMM) indicated individual prediction methods. G+C skew shown in the centre circle.

D.6 Merops compiled presence absence

Compiled presence and absence excel sheet for Merops output is attached under the file name Appendix D6- compiled MEROPs datasheet.xlsx. MEROPs outputs were included when score showed 60% identity to a matched peptidase. Isolate named across the columns where MEROPs matched ID is listed in rows.

Appendix E

Chapter 7

E.1 Putative SOS binding sit predictions

Annotated SOS binding site predictions made for isolates described in Chapter 7, section 7.2.1.1, are attached under the file name Appendix E1- putative SOS binding site annotations.xlsx. Excel sheets labelled with isolate name. Results indicate putative prediction of genes associated with an upstream SOS BOX.

E.2 Transconjugant plate assay data

Plate assay data is attached as an excel format file named Appendix E2- transconjugant plate assay data.xlsx. Data is shown as tabbed sheets for each assay protocol, with replicate data shown for each.

Appendix F

Chapter 8

F.1 Transcriptome data files

Transcriptome data output files are attached as three separate comparison files under the following file names:

- Appendix F2a- transcriptome data A1M02 v 477 + pA1M02.xlsx
- Appendix F1b- transcriptome data 477 v 477 + pAfpX.xlsx
- Appendix F1c- transcriptome data 477 v 477 + pA1M02.xlsx

F.2 pADAP annotation in comparison between A1M02 and 477 + pA1M02

The below table shows the pADAP specific annotations in A1M02 v 477 + pA1M02 comparisons, where AV7-9 are replicates of 477 + pA1M02, and MW7-9 are replicates of wildtype A1M02. Annotations were taken from pADAP GenBank accession AF135182.

Table F 1 Transcriptome data of pADAP genes in wildtype A1M02 (MW 7-9) and transconjugant 477 + pA1M02 (AV 7-9).

transcript_id	AV_7	AV_8	AV_9	MW_7	MW_8	MW_9	log2FoldChange	padj	pADAP annotation
JJLEJPGA_03734	189	318	286	202	266	158	-0.62719178	0.051413861	Hypothetical
JJLEJPGA_03735	101	133	145	90	100	76	-0.789310462	0.015969396	Tram family protein
JJLEJPGA_03736	1362	1515	1470	1651	1943	1980	0.06867111	0.716890194	SogL
JJLEJPGA_03737	30	40	39	35	47	46	-0.067338956	0.906183403	
JJLEJPGA_03738	45	36	30	45	58	74	0.362599995	0.479815957	IcmT trac
JJLEJPGA_03739	90	81	58	103	95	81	0.009970232	0.985862671	plasmid transfer ATPase TraJ
JJLEJPGA_03740	35	37	42	25	35	40	-0.494434936	0.327695349	type IV secretory system conjugative DNA transfer family protein
JJLEJPGA_03741	62	83	43	54	89	48	-0.294615351	0.576517127	DotD/TraH family lipoprotein
JJLEJPGA_03742	168	185	262	184	184	147	-0.521849132	0.111966793	shufflon system plasmid conjugative transfer pilus tip adhesin PilV
JJLEJPGA_03743	190	154	188	171	173	192	-0.272669925	0.375347952	prepilin peptidase
JJLEJPGA_03744	188	146	203	172	233	238	-0.04476587	0.891528665	lytic transglycosylase domain-containing protein
JJLEJPGA_03745	315	473	397	572	294	397	-0.119215514	0.810024764	PilS
JJLEJPGA_03746	330	351	418	375	530	520	0.071227242	0.757370549	type II secretion system F family protein
JJLEJPGA_03747	292	349	461	395	534	490	0.0700664	0.803257719	Flp pilus assembly complex ATPase component TadA
JJLEJPGA_03748	181	201	208	268	229	239	0.05778124	0.874483471	type IV pilus biogenesis protein PilP
JJLEJPGA_03749	188	284	296	214	313	237	-0.306185264	0.267101152	type 4b pilus protein PilO2
JJLEJPGA_03750	442	444	409	430	558	614	0.003544154	0.989222233	PilN family type IVB pilus formation outer membrane protein
JJLEJPGA_03751	224	180	247	179	157	147	-0.695026172	0.029244072	type IV pilus biogenesis protein PilM
JJLEJPGA_03752	278	337	290	233	295	319	-0.394113356	0.07581318	TcpQ domain-containing protein
JJLEJPGA_03753	50	93	71	76	68	70	-0.262751044	0.590881445	TcpQ domain-containing protein
JJLEJPGA_03754	546	436	434	411	453	529	-0.318106915	0.212850448	Hypothetical
JJLEJPGA_03755	1068	1068	1158	1733	1357	1570	0.24739094	0.386831138	Replication protein
JJLEJPGA_03756	6116	6601	6386	4695	7132	5942	-0.41530309	0.00036039	ATP-binding protein
JJLEJPGA_03757	1415	1548	1434	740	1143	1053	-0.896104883	5.9763E-11	SIR2 family protein
JJLEJPGA_03758	2229	2331	2123	1685	2390	2262	-0.38475113	0.00596506	GIY-YIG nuclease family protein
JJLEJPGA_03759	2025	2142	2081	1663	1982	1800	-0.488654282	0.001234822	ATP-dependent helicase
JJLEJPGA_03760	661	838	810	756	1182	930	0.003594164	0.987651887	Hypothetical
JJLEJPGA_03761	1470	1562	1701	852	1180	1214	-0.847935376	1.40971E-09	DNA polymerase V subunit UmuC
JJLEJPGA_03762	872	909	952	659	788	692	-0.643430854	7.82481E-05	ISKra4 family transposase
JJLEJPGA_03763	1728	2455	2528	1105	1579	1197	-1.087156021	2.43712E-09	ISKra4 family transposase
JJLEJPGA_03764	158	239	216	116	109	120	-1.099389846	0.000213792	plasmid pRiA4b ORF-3 family protein

JJLEJPGA_03765	327	473	388	228	383	344	-0.630560733	0.004741677	winged helix-turn-helix domain-containing protein
JJLEJPGA_03766	246	257	201	74	120	179	-1.241242546	0.000204301	Hypothetical
JJLEJPGA_03767	56	40	46	69	89	52	0.26611674	0.600525072	type 1 fimbrial protein
JJLEJPGA_03768	2	5	2	2	4	11	0.58851699	0.730034871	type 1 fimbrial protein
JJLEJPGA_03769	184	139	155	73	120	139	-0.850037497	0.006981887	fimbria/pilus outer membrane usher protein
JJLEJPGA_03770	17	18	11	19	13	11	-0.34904675	0.7040766	fimbria/pilus periplasmic chaperone
JJLEJPGA_03771	44	70	89	5	20	19	-2.547664148	4.59566E-06	type 1 fimbrial protein
JJLEJPGA_03772	131	84	98	68	125	169	-0.127426223	0.801808713	Fimbrial protein
JJLEJPGA_03773	329	243	269	215	380	333	-0.187277563	0.515659793	fimbria/pilus periplasmic chaperone
JJLEJPGA_03774	47	49	39	52	63	45	-0.048807061	0.930994857	Fimbrial protein
JJLEJPGA_03775	77	81	59	49	92	27	-0.696752826	0.202146048	Hypothetical
JJLEJPGA_03776	75	49	39	58	57	46	-0.30418374	0.592901955	Hypothetical
JJLEJPGA_03777	208	94	138	216	165	368	0.480227418	0.343317527	EAL domain-containing protein
JJLEJPGA_03778	155	132	130	62	152	179	-0.435987073	0.296537828	Hypothetical
JJLEJPGA_03779	143	144	124	45	23	40	-2.17404777	2.53117E-07	phage holin family protein
JJLEJPGA_03780	301	394	462	173	257	210	-1.155657152	1.31398E-07	Lysozyme
JJLEJPGA_03781	10501	12713	11177	6714	11172	7955	-0.728991669	1.84247E-06	insecticidal toxin complex protein A (SepA)
JJLEJPGA_03782	6929	8418	7300	4989	7100	4575	-0.743120355	7.64406E-05	SepB
JJLEJPGA_03783	1457	1773	1413	1212	1763	1426	-0.384176516	0.024645172	DUF2514 family protein
JJLEJPGA_03784	7439	8224	7390	6420	9814	7403	-0.274907025	0.06363763	SepC
JJLEJPGA_03785	673	710	617	851	1033	1026	0.245477393	0.192026861	IS91 family transposase
JJLEJPGA_03786	984	1123	1037	1241	1542	1399	0.118186648	0.506531013	IS91 family transposase
JJLEJPGA_03787	667	691	669	783	755	828	-0.051678957	0.848724574	IS91 family transposase
JJLEJPGA_03788	970	813	1108	653	885	945	-0.523419198	0.007906685	Hypothetical
JJLEJPGA_03789	273	322	356	482	530	446	0.338904346	0.15751626	IS66 family insertion sequence element accessory protein TnpB
JJLEJPGA_03790	46	43	85	54	75	25	-0.46078343	0.472107462	IS66 family transposase
JJLEJPGA_03791	16	24	29	0	9	8	-2.404250376	0.013031287	Lysozyme
JJLEJPGA_03792	111	139	109	71	99	73	-0.865786662	0.004824915	Afp 1
JJLEJPGA_03793	2380	2442	2794	571	900	767	-2.078764086	1.78842E-59	Afp 2
JJLEJPGA_03794	69030	97702	80655	876	1303	1460	-6.396316099	4.7064E-296	Afp 3
JJLEJPGA_03795	42492	67777	53322	495	817	425	-6.866104132	3.4437E-195	Afp 4
JJLEJPGA_03796	5193	8037	5960	158	166	189	-5.507587032	1.3418E-154	Afp 5
JJLEJPGA_03797	1025	1790	1223	7	9	1	-8.189541821	1.22726E-76	Afp 6
JJLEJPGA_03798	8345	15533	9548	97	99	41	-7.401351171	6.3664E-104	Afp 7
JJLEJPGA_03799	20704	41653	24856	214	315	299	-7.02237569	5.1441E-181	Afp 8

JJLEJPGA_03800	11424	19219	13775	116	107	115	-7.308558306	1.6795E-213	Afp 9
JJLEJPGA_03801	15157	22540	15666	283	513	460	-5.732427457	3.3146E-187	Afp 10
JJLEJPGA_03802	44637	77968	59394	399	436	286	-7.612442867	4.87E-207	Afp 11
JJLEJPGA_03803	66142	104642	78506	1433	2091	1624	-5.89962624	1.3708E-234	Afp 12
JJLEJPGA_03712	69	53	95	35	47	63	-0.884934372	0.045680892	Afp 13
JJLEJPGA_03713	38	23	38	24	25	32	-0.581353172	0.328218936	Afp 14
JJLEJPGA_03714	199	133	171	95	126	71	-1.085569028	0.001892728	Afp 15
JJLEJPGA_03715	299	173	212	142	215	191	-0.641230093	0.044086581	Afp 16
JJLEJPGA_03716	2306	1760	1568	948	1302	1078	-1.070822776	7.17056E-07	Afp 17
JJLEJPGA_03717	9646	9972	8986	5033	10652	7406	-0.649893919	0.000783573	Afp 18
JJLEJPGA_03718	2455	2799	3319	3218	4523	4170	0.174016348	0.291724325	Hypothetical
JJLEJPGA_03719	1215	1204	1397	1109	1922	1638	-0.030007411	0.88233634	DUF1669 domain-containing protein
JJLEJPGA_03720	749	681	783	828	969	913	0.002202351	0.99218185	Conjugal transfer protein TrbC
JJLEJPGA_03721	229	187	199	239	178	146	-0.374036776	0.358417533	Hypothetical
JJLEJPGA_03722	268	367	259	356	317	483	0.095830372	0.801991834	Conjugal transfer protein TrbA
JJLEJPGA_03723	15979	14386	15779	5204	12160	8732	-1.171379537	1.29455E-08	Hypothetical
JJLEJPGA_03724	1010	821	876	603	750	740	-0.671592793	0.000374563	DotA/TraY family protein
JJLEJPGA_03725	120	135	166	93	129	112	-0.633478701	0.022139638	conjugal transfer protein TraX
JJLEJPGA_03726	167	289	269	109	270	183	-0.706068393	0.035932188	conjugal transfer protein TraW
JJLEJPGA_03727	31	42	47	11	33	24	-1.162201495	0.036232638	Hypothetical
JJLEJPGA_03728	693	789	891	743	1015	826	-0.17453828	0.343717346	conjugal transfer protein TraU
JJLEJPGA_03729	103	173	174	68	97	50	-1.357252797	0.000269417	conjugal transfer protein TraT
JJLEJPGA_03730	140	249	197	227	274	213	0.002465739	0.994596665	Hypothetical
JJLEJPGA_03731	156	122	110	162	235	213	0.337577594	0.269933741	conjugal transfer protein TraQ
JJLEJPGA_03732	115	129	125	183	241	200	0.45979537	0.065174746	Hypothetical
JJLEJPGA_03733	37	46	53	93	63	108	0.705085012	0.13840087	conjugal transfer protein TraO
JJLEJPGA_03734	189	318	286	202	266	158	-0.62719178	0.051413861	Hypothetical